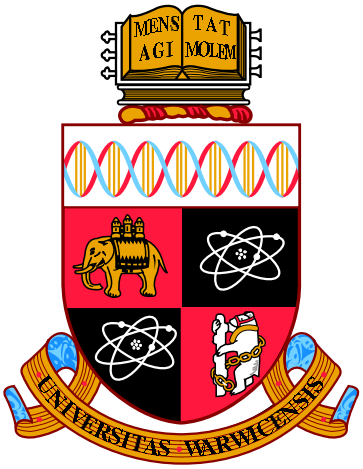


Localisation and Symmetry in Computational Pathology

by

Simon Graham

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Computer Science

Mathematics for Real-World Systems

September 2020

Contents

List of Tables

List of Figures

Acknowledgments

I would like to give special thanks to my supervisor Prof Nasir Rajpoot, who has helped mould me into a capable researcher in the area of computational pathology. His guidance, support and friendship will not be forgotten and I am extremely thankful for all of the opportunities that he has provided during my PhD.

Thank you to Prof David Epstein for our stimulating discussions, especially over the past year, where we spent countless hours discussing various mathematical concepts. He taught me to always try and fully understand *why* something works and to pay special attention to mathematical notation.

I would like to thank all of the current and previous members of the Tissue Image Analytics (TIA) Lab at the University of Warwick: Dr Talha Qaiser, Muhammad Shaban, Navid Alemi Koohbanani, Dr Najah Alsubaie, Mary Shapcott, Ruqayya Awan, Saad Bashir, Jevgenij Gamper, Dr Sajid Javed, Dr Moazam Fraz, Hammam Alghamdi, John Pocock, Dr Nima Hatami, Dr Mohsin Bilal, Sirijay Deshpande and Rawan Albusayli. I personally thank Dr Talha Qaiser for his patience and guidance when I first joined the lab as a masters student. Without this, I may never have gone on to pursue a PhD in this area. In addition, I must express my thanks to the members of PathLAKE including: Dr Shan Raza, Dr Fayyaz Minhas, Dr Hadi Saki, Dr Wenqi Lu, Dr Noorul Wahab and Dr Young Park. I have enjoyed working with you all and look forward to our future collaborations on the PathLAKE project.

My external collaborators have played an integral part in the completion of my thesis. I thank my clinical collaborators who have helped develop my understanding of histopathology. Specifically, I thank Prof David Snead, Dr Yee Wah

Tsang, Dr Ayesha Azam, Dr Ali Khurram, Dr Ksenija Benes and Dr Katherine Hewitt. I also thank Dr Hao Chen and Prof Pheng-Ann Heng from the Chinese University of Hong Kong, along with Dang Vu and Dr Jin Tae Kwak from Sejong University.

I thank my mother and sister, Jackie and Charlotte, for their support and encouragement during my PhD. I thank my mother for her unconditional love and teaching me strength and perseverance in all aspects of life. A special thanks must go to my girlfriend Elysa who has shown her continual love and support over the last few years. Her ability to make me laugh and to pick me up when times are tough has pushed me to keep going and to remain positive during the course of my PhD.

I acknowledge the financial support from the Engineering and Physical Sciences Research Council (EPSRC) and Medical Research Council (MRC), provided as part of the Mathematics for Real-World Systems CDT.

I dedicate this thesis to my father, David, who passed away due to cancer in 2013.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. I declare that, except where acknowledged, the material presented in this thesis is my own work, and has not been previously submitted for obtaining an academic degree.

Simon Graham
20th September 2020

List of Publications

First-Authored Publications

Journal Articles

- **Simon Graham**, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak and Nasir Rajpoot. HoVer-Net: Simultaneous Segmentation and Classification of Nuclei in Multi-Tissue Histology Images. *Medical Image Analysis, 58, p. 101563 (2019).* **[Chapter 3]**

- **Simon Graham**, Hao Chen, Jevgenij Gamper, Qi Dou, Pheng-Ann Heng, David Snead, Yee Wah Tsang and Nasir Rajpoot. MILD-Net: Minimal Information Loss Dilated Network for Gland Instance Segmentation in Colon Histology Images. *Medical Image Analysis, 52, pp. 199-211 (2019).* **[Chapter 4]**

- **Simon Graham**, David Epstein and Nasir Rajpoot. Dense Steerable Filter CNNs for Exploiting Rotational Symmetry in Histology Images. *IEEE Transactions on Medical Imaging, in press (2020).* **[Chapter 5]**

Conference Papers

- **Simon Graham** and Nasir Rajpoot. Sams-net: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images. *IEEE 15th International Symposium on Biomedical Imaging, pp. 590-594 (2018).*

- **Simon Graham**, Muhammad Shaban, Talha Qaiser, Navid Alemi Koohbanani, Syed Ali Khurram and Nasir Rajpoot. Classification of lung cancer histology

images using patch-level summary statistics. *Medical Imaging 2018: Digital Pathology, Vol. 10581, p. 1058119. International Society for Optics and Photonics. (2018).* **[Chapter 2]**

- **Simon Graham**, Hao Chen, Qi Dou, Pheng-Ann Heng and Nasir Rajpoot. MILD-Net: Minimal Information Loss Dilated Network for Gland Instance Segmentation in Colon Histology Images. *Medical Imaging with Deep Learning (2018).* **[Chapter 4]**

- **Simon Graham**, David Epstein and Nasir Rajpoot. Rota-Net: Rotation Equivariant Network for Simultaneous Gland and Lumen Segmentation in Colon Histology Images. *European Congress on Digital Pathology, pp. 109-116 (2019).* **[Chapter 5]**

Co-Authored Publications

Journal Articles

- Quoc Dang Vu, **Simon Graham**, Tahsin Kurc, Minh Nguyen Nhat To, Muhammad Shaban, Talha Qaiser, Navid Alemi Koohbanani, Syed Ali Khurram, Jayashree Kalpathy-Cramer, Tianhao Zhao, Rajarsi Gupta, Jin Tae Kwak, Nasir Rajpoot, Joel Saltz, Keyvan Farahani. Methods for Segmentation and Classification of Digital Microscopy Tissue Images. *Frontiers in Bioengineering and Biotechnology, 7 (2019).*

- Huangjing Lin, Hao Chen, **Simon Graham**, Qi Dou, Nasir Rajpoot and Pheng-Ann Heng. Fast ScanNet: Fast and Dense Analysis of Multi-Gigapixel Whole-Slide Images for Cancer Metastasis Detection. *IEEE Transactions on Medical Imaging, 38(8), pp. 1948-1958 (2019).*

- Muhammad Moazam Fraz, Syed Ali Khurram, **Simon Graham**, Muhammad Shaban, Maryam Hassan, Asif Loya, Nasir Rajpoot. FABnet: Feature Attention Based Network for Simu ltaneous Segmentation of Microvessels and

Nerves in Routine Histology Images of Oral Cancer. *Neural Computing and Applications, pp. 1-14 (2019).*

- Shan E Ahmed Raza, Linda Cheung, Muhammad Shaban, **Simon Graham**, David Epstein, Stella Pelengaris, Michael Khan, Nasir Rajpoot. Micro-Net: A Unified Model for Segmentation of Various Objects in Microscopy Images. *Medical Image Analysis, 52, pp. 160-173 (2019).*

- Mitko Veta, Yujing J Heng, Nikolas Stathonikos, Babak Ehteshami Bejnordi, Francisco Beca, Thomas Wollmann, Karl Rohr, Manan A Shah, Dayong Wang, Mikael Rousson, Martin Hedlund, David Tellez, Francesco Ciompi, Erwan Zerhouni, David Lanyi, Matheus Viana, Vassili Kovalev, Vitali Liauchuk, Hady Ahmady Phoulady, Talha Qaiser, **Simon Graham**, Nasir Rajpoot, Erik Sjöblom, Jesper Molin, Kyunghyun Paeng, Sangheum Hwang, Sunggyun Park, Zhipeng Jia, I Eric, Chao Chang, Yan Xu, Andrew H Beck, Paul J van Diest, Josien PW Pluim. Predicting Breast Tumor Proliferation from Whole-Slide Images: the TUPAC16 Challenge. *Medical image analysis, 54, pp. 111-121 (2019).*

- Neeraj Kumar, Ruchika Verma, Deepak Anand, Yanning Zhou, Omer Fahri Onder, Efstratios Tsougenis, Hao Chen, Pheng Ann Heng, Jiahui Li, Zhiqiang Hu, Yunzhi Wang, Navid Alemi Koohbanani, Mostafa Jahanifar, Neda Zamani Tajeddin, Ali Gooya, Nasir Rajpoot, Xuhua Ren, Sihang Zhou, Qian Wang, Dinggang Shen, Cheng Kun Yang, Chi Hung Weng, Wei Hsiang Yu, Chao Yuan Yeh, Shuang Yang, Shuoyu Xu, Pak Hei Yeung, Peng Sun, Amirreza Mahbod, Gerald Schaefer, Isabella Ellinger, Rupert Ecker, Orjan Smedby, Chunliang Wang, Benjamin Chidester, That Vinh Ton, Minh-Triet Tran, Jian Ma, Minh N Do, **Simon Graham**, Quoc Dang Vu, Jin Tae Kwak, Akshaykumar Gunda, Raviteja Chunduri, Corey Hu, Xiaoyang Zhou, Dariush Lotfi, Reza Safdari, Antanas Kascenas, Alison O'Neil, Dennis Eschweiler, Johannes Stegmaier, Yanping Cui, Baocai Yin, Kailin Chen, Xinmei Tian, Philipp Gruening, Erhardt Barth, Elad Arbel, Itay Remer, Amir Ben-Dor, Ekaterina

Sirazitdinova, Matthias Kohl, Stefan Braunewell, Yuexiang Li, Xinpeng Xie, Linlin Shen, Jun Ma, Krishanu Das Baksi, Mohammad Azam Khan, Jaegul Choo, Adriá n Colomer, Valery Naranjo, Linmin Pei, Khan M Iftekharuddin, Kaushiki Roy, Debotosh Bhattacharjee, Anibal Pedraza, Maria Gloria Bueno, Sabarinathan Devanathan, Saravanan Radhakrishnan, Praveen Koduganty, Zihan Wu, Guanyu Cai, Xiaojie Liu, Yuqin Wang, Amit Sethi. A Multi-Organ Nucleus Segmentation Challenge. *IEEE Transactions on Medical Imaging (2019).*

- Jevgenij Gamper, Navid Alemi Koohbanani, **Simon Graham**, Mostafa Jahanifar, Syed Ali Khurram, Ayesha Azam, Katherine Hewitt, Nasir Rajpoot. PanNuke: Pan-Cancer Multi-Organ Dataset for Nuclear Segmentation and Classification. *Submitted to Nature Scientific Data (2020).* [**Appendix A**]

Conference Papers

- Saad Ullah Akram, Talha Qaiser, **Simon Graham**, Juho Kannala, Janne Heikkilä and Nasir Rajpoot. Leveraging Unlabeled Whole-Slide-Images for Mitosis Detection. *Computational Pathology and Ophthalmic Medical Image Analysis, pp. 69-77 (2018).*

- Muhammad Moazam Fraz, Muhammad Shaban, **Simon Graham**, Syed Ali Khurram and Nasir Rajpoot. Uncertainty Driven Pooling Network for Microvessel Segmentation in Routine Histology Images. *Computational Pathology and Ophthalmic Medical Image Analysis, pp. 156-164 (2018).*

- Yanning Zhou, **Simon Graham**, Navid Alemi Koohbanani, Muhammad Shaban, Pheng-Ann Heng and Nasir Rajpoot. CGC-Net: Cell Graph Convolutional Network for Grading of Colorectal Cancer Histology Images. *IEEE International Conference on Computer Vision Workshops (2019).*

Abstract

Conventional assessment of Haematoxylin and Eosin (H&E) stained tissue slides is performed via visual examination under the microscope by a pathologist and often serves as the *gold standard* in cancer diagnosis. Standard diagnostic practice requires pathologists to follow a descriptive set of guidelines and is, therefore, prone to suffer from inter-observer variability due to differences in interpretation of histological patterns. Furthermore, each tissue slide may contain tens of thousands of cells and, therefore, accurate quantification and morphological analysis of the tissue in the entire slide is not feasible. Recently, there has been a growing trend towards a digital pathology workflow, where tissue slides are digitised with a high-resolution scanner to obtain Whole-Slide Images (WSIs). This enables the development of automatic tools that can objectively analyse and quantify the vast amount of pixel information contained in multi-gigapixel WSIs.

In this thesis, we initially introduce the challenge of analysing large-scale WSIs for histology image analysis by presenting a preliminary WSI classification framework. Here, we predict the diagnosis of a slide by: (i) dividing the WSI into small image regions (patches), (ii) making predictions independently on each patch and then (iii) predicting the overall slide diagnosis by aggregating patch-level results.

In the remainder of the thesis, we focus on developing automated methods that localise objects and structures of interest in the tissue and that leverage the presence of rotational symmetry in histology images. Localisation of nuclei and other components, such as glands, allows further exploration of digital biomarkers and serves as a fundamental pre-requisite for downstream analysis. On the other hand,

exploitation of rotational symmetry for histology image analysis enables models to be tailored to the specific geometry of microscopy images, where there exists no underlying global orientation.

In this regard, we present the first single convolutional neural network (CNN) for simultaneous segmentation and classification of nuclei. The CNN uses the concept of *horizontal and vertical maps* to separate clustered nuclei and utilises a devoted upsampling branch to accurately perform nuclear classification. We then propose a novel CNN for gland segmentation that counters the loss of information caused by max-pooling by reintroducing the original image at multiple points within the network. To enable localisation of glands with varying size, we additionally incorporate *atrous* spatial pyramid pooling.

To leverage the prior knowledge that histology images are symmetric under rotation, it is desirable for CNNs to be rotation-equivariant. This guarantees that features transform as expected with rotation of the input. In this thesis, we perform the first thorough analysis of various rotation-equivariant models for histology image analysis. We then develop a CNN for simultaneous segmentation of glands and lumen that achieves rotation-equivariance by using group-convolutions with multiple rotated copies of each filter. Finally, we propose a general CNN for histology image analysis that employs the concept of group-convolution and defines filters as a linear combination of steerable basis filters. This enables exact rotation and decreases the number of trainable parameters compared to standard filters.

Abbreviations

AJI: Aggregated Jaccard Index

ASPP: Atrous Spatial Pyramid Pooling

AUC: Area Under the Receiver Operating Characteristic Curve

BCE: Binary Cross Entropy

CRC: Colorectal Adenocarcinoma

CRC: Colorectal Cancer

CPath: Computational Pathology

CNN: Convolutional Neural Network

DL: Deep Learning

DSF: Dense Steerable Filter Network

DQ: Detection Quality

DPath: Digital Pathology

FCN: Fully Convolutional Network

GCN: Graph Convolutional Network

GPU: Graphics Processing Unit

GT: Ground Truth

G-CNN: Group Equivariant Convolutional Neural Network

H&E: Haematoxylin & Eosin

HPF: High Power Field

IHC: Immunohistochemistry

LUAD: Lung Adenocarcinoma

LUSC: Lung Squamous Cell Carcinoma

ML: Machine Learning

MV: Majority Voting

MSE: Mean Squared Error

MIL: Minimal Information Loss

ND: Non-Diagnostic

NSCLC: Non-Small Cell Lung Cancer

NC: Nuclear Classification

NP: Nuclear Pixel

PQ: Panoptic Quality

RF: Random Forest

RTS: Random Transformation Sampling

SQ: Segmentation Quality

TAMs: Tumour Associated Macrophages

TCGA: The Cancer Genome Atlas

TILs: Tumour Infiltrating Lymphocytes

TNBC: Triple Negative Breast Cancer

TTA: Test-Time Augmentation

UHCW: University Hospitals Coventry and Warwickshire

VF-CNN: Vector Field Convolutional Neural Network

WSI: Whole-Slide Image

Chapter 1

Introduction

1.1 Cancer

Cancer is the broad term for a group of diseases that describe the over-proliferation of cells and is responsible for an estimated 10 million global deaths per year [6]. Cells are the fundamental buildings blocks of the body and constantly divide to enable growth and repair. However, as a result of the interaction of multiple genetic and environmental factors, this cell division may become uncontrolled, leading to an abnormal mass of cells forming a tumour. Benign tumours describe an area of abnormal cell growth, but are generally harmless unless it is pressing on nearby tissues, nerves, or blood vessels [152]. On the other hand, cancerous (or malignant) tumours invade the nearby tissue, breaking through the basal lamina that define the tissue boundaries [102], and spread to other organs in the body. This spread of tumour cells to secondary areas of growth is referred to as metastasis and its extent is referred to as the stage of the cancer. The cancer grade is a description based on the appearance of tumour cells, where a high grade implies that tumour cells have lost their typical cellular characteristics.

In addition to the stage and grade of cancer, the complex interaction of various cells within the tumour microenvironment (TME) provide insight into cancer development. For example, the spatial arrangement of tumour infiltrating lymphocytes (TILs) is associated with clinical outcome in several cancers [48] and tumour associated macrophages (TAMs) influence multiple diverse processes in various tumours [122]. Therefore, a thorough analysis of the tissue and the TME is essential to determine the appropriate treatment for each patient.

1.2 Histological Analysis

1.2.1 Slide Preparation

Typically, cancer diagnosis is performed via visual examination of histological tissue sections under the microscope by a pathologist and involves analysing both cell-level information and the tissue architecture. Before this examination can take place, the tissue must be appropriately prepared. This preparation consists of the following steps: (i) preserving the tissue using fixation; (ii) embedding the tissue in a paraffin block; (iii) cutting the paraffin block into thin sections (3-5μm); (iv) mounting the sections on glass slides and finally (v) staining mounted tissue sections to highlight important components. Haematoxylin and Eosin (H&E) are the most commonly used stains for morphological analysis of the tissue. Haematoxylin binds to the DNA and stains the nuclei dark blue/purple, whereas Eosin stains the extracellular matrix and cytoplasm pink. Other staining techniques such as Immunohistochemistry (IHC) are often used to detect the presence of specific protein markers. However, in this thesis, we limit our analysis to H&E slides.

As part of the preparation process, there can be large variation in the appearance between different stained tissue samples. For example, thicker specimens tend to stain the tissue darker and differences in the temperature, stain concentration and duration of staining can also lead to variation. As well as this, there may exist artefacts in the prepared tissue, including: tissue folds and regions with tissue scoring that result from cutting sections with a blunt blade. It is common for such artefacts to appear, but must not impact the pathologist's ability to diagnose a slide. On the top row of Figure 1.1 we show an example two tissue regions stained with Haematoxylin and Eosin, yet their visual appearance is strikingly different. On the bottom row, we display an example of tissue scoring and tissue folds that may be introduced as part of the standard preparation process.

1.2.2 Histological Types

Cancers are diagnosed according to the tissue type in which cancer originates from (histological type) and the organ where the cancer first developed (primary site). In terms of categorising cancer by its histology, there are hundreds of different types. However, they can be broadly grouped into the following categories [5]:

- Carcinoma
- Sarcoma

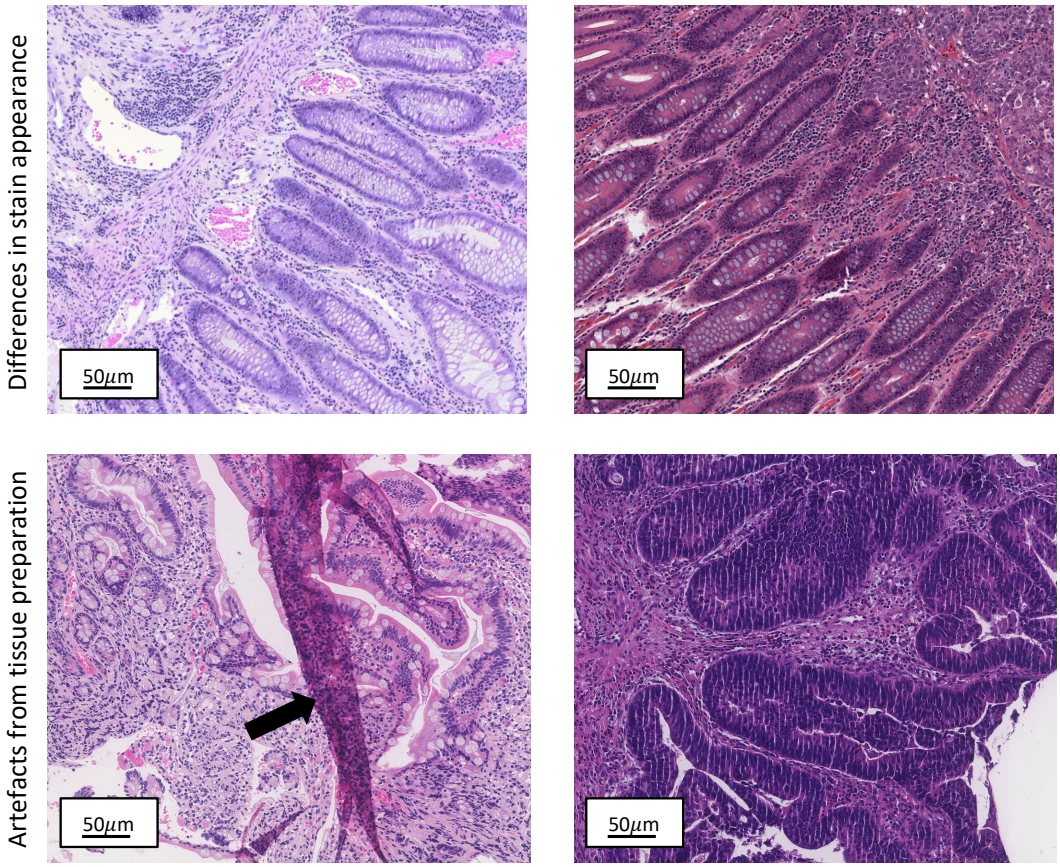


Figure 1.1: Image regions from H&E stained tissue, highlighting stain variability and displaying artefacts from tissue preparation. On the bottom row, we display an image with a tissue fold (left) indicated by the black arrow and an image with tissue scoring (right).

- Leukemia
- Lymphoma
- Myeloma
- Mixed

Carcinoma is a cancer type that develops in the epithelial cells that line the organs in the body and accounts for around 80 to 90% of all cancer cases [2]. The two most common sub-types of carcinoma are adenocarcinoma and squamous cell carcinoma. Adenocarcinoma describes a cancer that forms in mucus-secreting glands, whereas squamous cell carcinoma originates in the squamous cells that line the tissue. Sarcoma is a cancer that starts in the connective tissues, including the

bone, cartilage, muscle and blood vessels and often closely resembles the tissues in which they grow. Leukemia originates in the blood forming tissue such as the bone marrow and causes abnormal blood cells to be produced and go into the blood. Lymphoma and myeloma are cancers that start in the immune cells. Specifically, lymphoma starts in the glands or nodes of the lymphatic system, whereas myeloma starts in the plasma cells of the bone marrow. Finally, mixed cancer types are a combination of histological types that can be between categories or within a single category. In this thesis, our analysis is focused on carcinomas, where they typically originate in secretory organs such as the breast, colon, lungs, bladder or prostate. As mentioned in Section 1.2.1, diagnosis of the cancer type is is done via histological examination of the tissue under the microscope.

We study a range of cancers in this thesis, but two of the most extensively studied types are colorectal cancer (CRC) and lung cancer. Below we give a general overview of these cancer types and provide a description of some common histological characteristics.

1.2.3 Colorectal Cancer

There are approximately 16,300 CRC deaths in the UK every year (2015-2017) [1], making it the UK's second leading cause of cancer death. CRC is the fourth most commonly occurring cancer in the United Kingdom (UK), where in 2017 it accounted for around 11% of all new cancer cases. CRC is the general term that combines both colon and rectal cancers and is part of the final stages of digestive system. The colon is responsible for processing indigestible food material after most of the nutrients have been absorbed in the small intestine. This material is then passed to the rectum and then leaves the body via the anus. In order to ease the transportation of waste material through the digestive system, the colon and rectum possess a network of mucus-secreting glands that project from the inner surface of the colon to the underlying connective tissue, as seen in Figure 1.2. The appearance of the glands is also determined by how the glands are cut. For example, in Figure 1.2 if the gland marked by *2* was to be cut in the direction of the black dashed line, then it would appear like the gland marked by *3*.

The most common form of CRC is CRA, where it accounts for around 95% of all cases [46]. CRA starts in the cells lining the glands of the colon wall and therefore the degree of glandular formation serves as the basis for histological tumour grading. In well differentiated CRA, over 95% of tumours are gland forming, whereas in moderately and poorly differentiated CRA there are significantly less gland forming tumours. In practice, most CRAs are diagnosed as moderately dif-

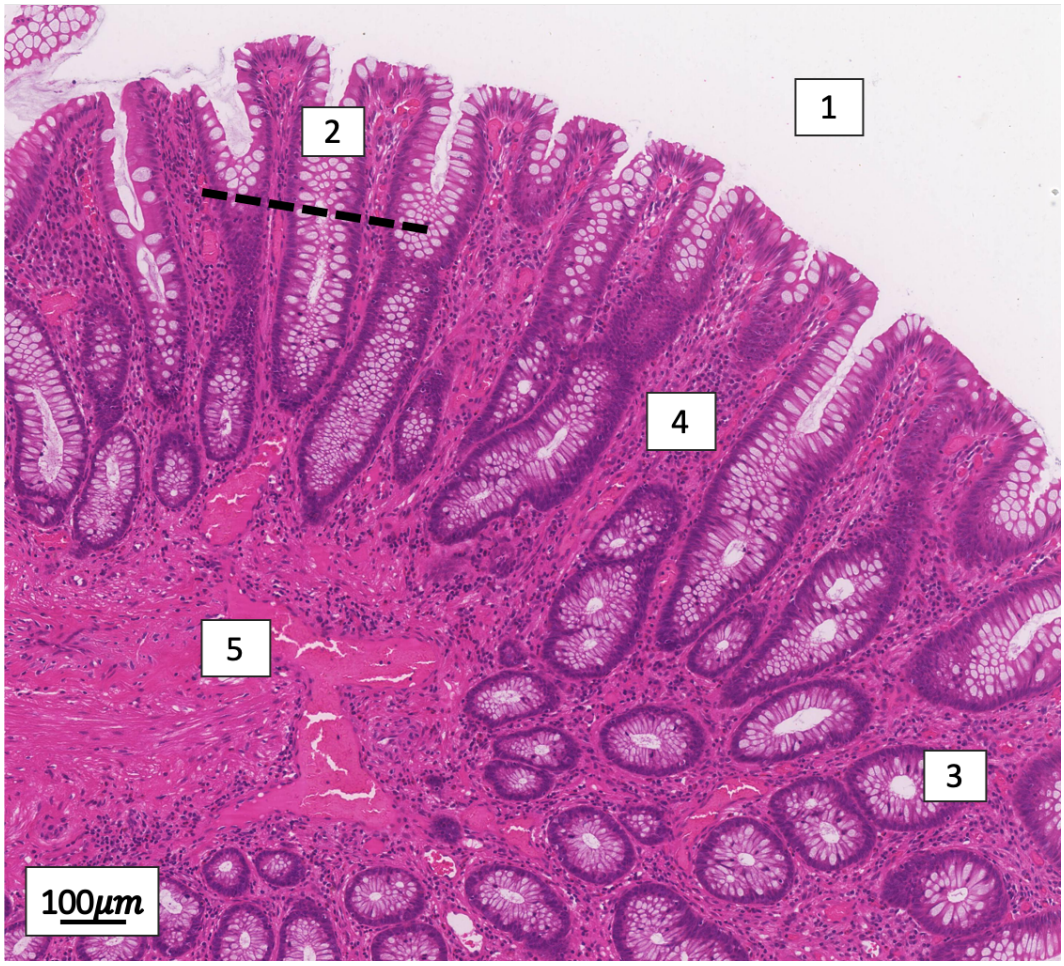


Figure 1.2: Image region from a colon H&E stained tissue section. 1) lumen, 2) intestinal gland (crypt), 3) glands when cut in the direction of black line, 4) lamina propria, 5) mucosa.

ferentiated (around 70%), whereas well and poorly differentiated CRAs account for around 10% and 20%, respectively [46]. In Figure 1.3 we display a selection of image regions extracted from a series of WSIs. Here, we can see that as the grade of CRA increases, typical glandular appearance is less evident.

1.2.4 Lung Cancer

Lung cancer is the leading cause of cancer related death in the UK, where it accounted for around 35,300 lung cancer deaths pe year during 2015-2017 [3]. The lungs are a major component of the respiratory system, where they are responsible for the process of gas exchange. Here, air enters the body via the trachea which then splits into two bronchi. One bronchus enters each lung and then further separates

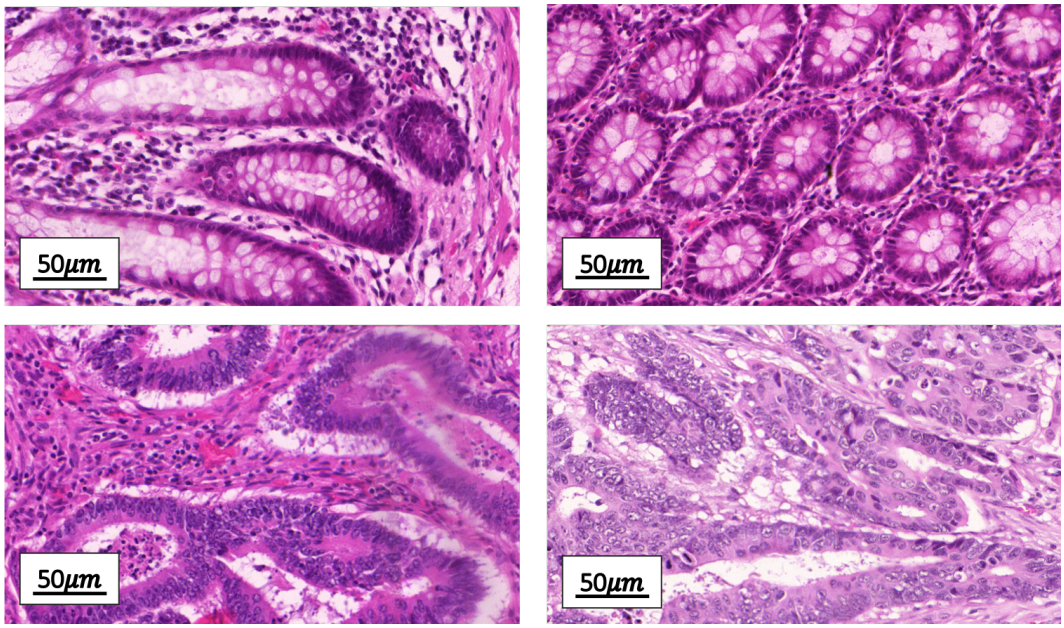


Figure 1.3: Images taken from H&E tissue sections showing the loss of glandular formation with increasing grade of cancer. Top row: normal glands, bottom row: moderately and poorly differentiated glands.

into around 30,000 smaller tubes, named bronchioles. At the end of each bronchiole exists a cluster of air sacs called alveoli that exchange oxygen and carbon dioxide molecules to and from the bloodstream. In total, lungs contain around 600 million alveoli. In Figure 1.4 we show an image region of a tissue section taken from the lung, where we can see normal bronchioles and alveoli. Note, the alveoli have a thin wall to enable efficient diffusion.

There are two main types of lung cancer: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). NSCLC accounts for over 80% of cases, where the two major histological types are lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) [42]. LUAD accounts for about 40% of all lung cancers and originates in the mucus-secreting glands within the lung. LUAD is histologically heterogeneous, where there exists 5 distinct growth patterns [138, 124, 144] which characterise the architecture of the tumour. The 5 growth patterns that exist in the lung are: acinar, papillary, micro-papillary, lepidic and solid. However, over 80% of LUAD cases are diagnosed as a mixed sub-type, consisting of two or more growth patterns. LUSC accounts for about 25-30% of all lung cancers and originates in the tissue that lines the air passages within the lung. In well differentiated LUSC, typical features include keratinisation, often in pearl formation, and inter-cellular bridging. In Figure 1.5, we show some image regions from LUAD and LUSC WSIs

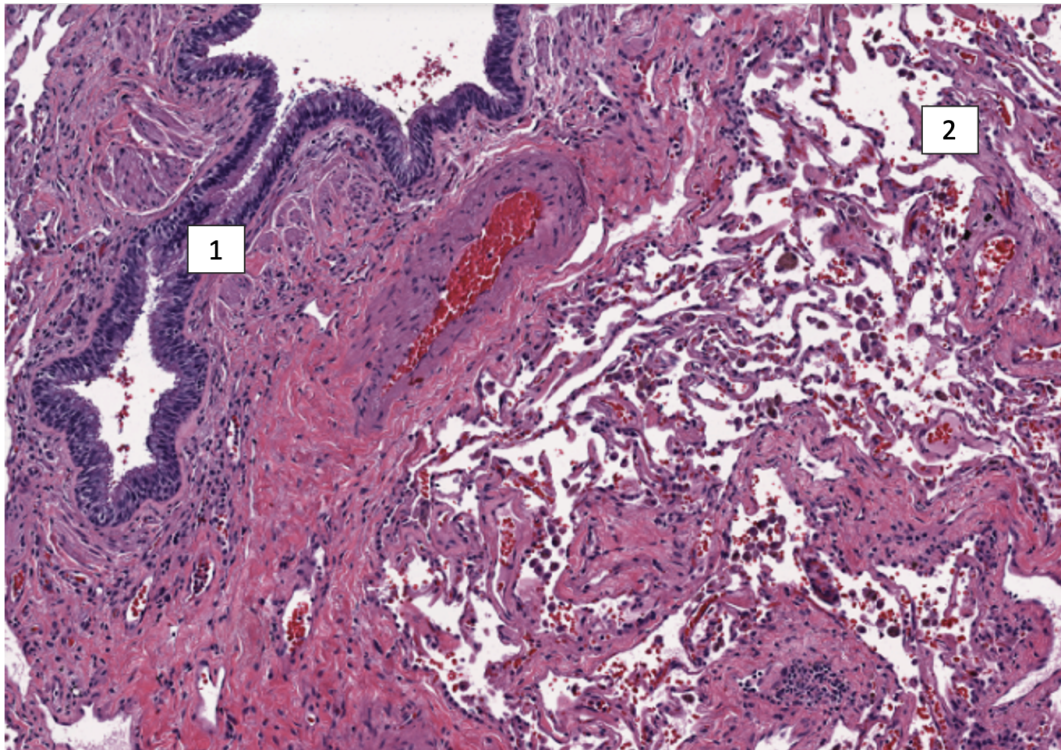


Figure 1.4: Image of H&E tissue sections from lung. 1) bronchiole, 2) alveoli.

that display typical histological characteristics. It is important for a pathologist to be able to differentiate between these two cancer types because several therapies are now available only for LUAD and certain specific LUAD mutations [141].

1.2.5 Challenges with Visual Examination

Visual examination of histology slides is a laborious and potentially time-consuming task because pathologists need to thoroughly inspect each case to ensure an accurate diagnosis. In the case of biopsy screening via histological examination, thousands of cases in many hospitals need to be diagnosed per year and therefore a quick turnaround time for the slides is essential. This poses a key challenge, especially when most NHS histopathology departments don't have enough staff to meet clinical demand [4]. Furthermore, there is often significant variability in the diagnosis given between different pathologists [127, 115, 54]. For example, certain cancer grading guidelines, such as the Gleason grading system [45] for prostate and the Scarff-Bloom Richardson grading system for breast [44], rely on the pathologist's interpretation of the tissue appearance. This interpretation is inherently subjective, which leads to differences in diagnosis. For example, one component of the Scarff-Bloom Richard-

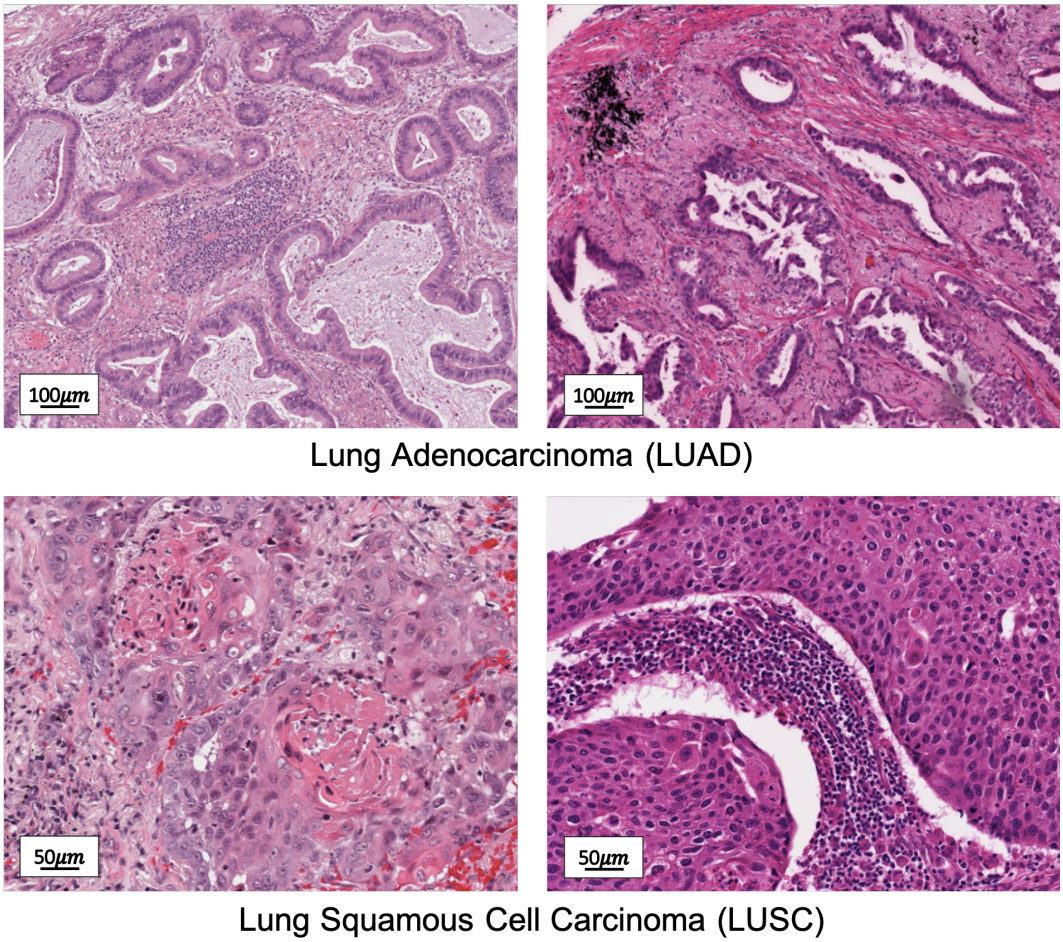


Figure 1.5: Example image regions from lung adenocarcinoma and lung squamous cell carcinoma tissue sections.

son grading system is counting the number of mitotic cells (cells undergoing division) in 10 regions displaying high proliferative activity. The selection of these 10 regions will differ between pathologists, which will inevitably lead to disagreement over the final count. Also, this task is very labour-intensive and mitotic figures can be easily missed when a pathologist has many slides to analyse. More generally, less experienced pathologists usually display variability in the diagnosis [38, 84] and there is often low agreement between pathologists when presented with a rare cancer type [84]. As a result of the aforementioned challenges, it is clear that there is a need for a more objective measure of histopathology slides that can also help to reduce the workload of the pathologist.

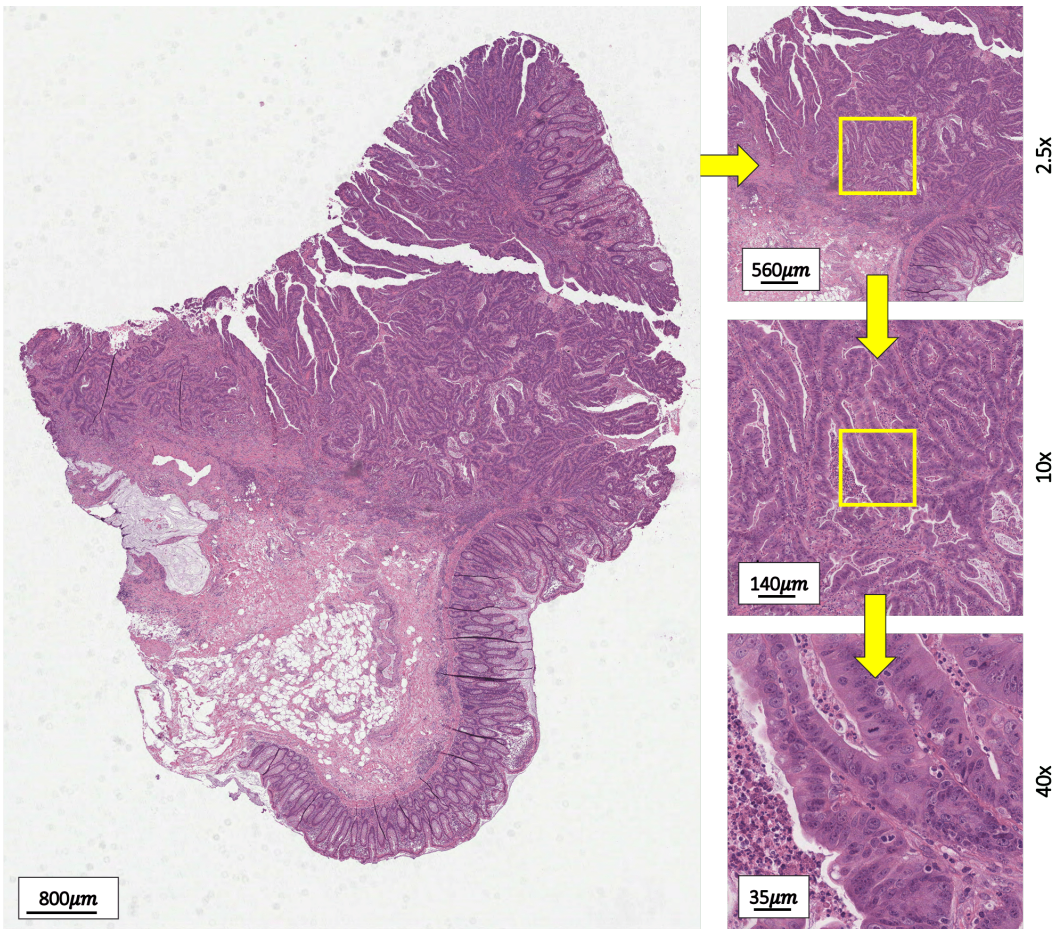


Figure 1.6: Example whole-slide image of colorectal tissue highlighting the multi-resolution structure.

1.3 Digital and Computational Pathology

1.3.1 Whole-Slide Images

Since the advent of digital slide scanners, tissue slides can now be converted into digital images that allow the reproduction of the original slide on a computer workstation [137]. These digital sides are often referred to as whole-slide images (WSIs) and are typically stored in a pyramid format, where each level of the pyramid represents a different magnification level (Figure 1.6). The highest magnification level is commonly at 40× (~0.25μm/pixel scan resolution), which approximately results in a 56GB image containing around 20 billion pixels [35]. Due to this huge file size, compression formats such as JPEG2000 and JPEG are often used to significantly reduce the size. Even with compression, reading the WSIs is a challenge because

standard image libraries are built in such a way that the entire image is uncompressed and loaded into memory. Instead, image libraries such as OpenSlide [56] are built for efficient data retrieval from WSIs of multiple different file formats.

1.3.2 Computational Pathology

One major advantage of acquiring WSIs is that it presents the opportunity for the development of computational algorithms to automatically analyse the tissue characteristics of each slide, which can help to overcome some of the challenges mentioned in Section 1.2.5. The study of such tools for pre-processing and subsequent analysis of WSIs is referred to as Computational Pathology (CPath). In general, application of CPath can be categorised into the following groups: (i) pre-processing, (ii) detection and segmentation, (iii) cancer type and grade prediction and (iv) prediction of prognosis. Below we provide some specific examples within each category.

Pre-processing: As mentioned in Section 1.2.1, there can be a significant difference in the colour appearance between different WSIs, due to variation in tissue preparation. Furthermore, the optics, image acquisition device and image acquisition algorithm used by different slide scanners can play a role in how the colour of the images appear on the computer monitor [162]. Despite the fact that pathologists can still diagnose tissue slides successfully in the presence of stain variation, the performance of CPath algorithms may be negatively affected. Therefore, algorithms can be developed to standardise the stain appearance between digital images before subsequent analysis [145, 81, 108]. As well as this, there may be artefacts present within each WSI such as tissue folds, ink markings and out-of-focus regions. Pre-processing algorithms that detect these artefacts may help inform whether a glass slide needs re-scanning or may be used to focus the analysis within artefact-free areas.

Detection and segmentation: WSIs contain a huge amount of pixel information that a pathologist needs to decipher to reach a diagnosis. Computational algorithms can assist with the detection, quantification and localisation of components within the tissue and can therefore help increase diagnostic accuracy and reduce the time a pathologist needs to spend on each slide. In particular, CPath allows WSI nuclei quantification [134] that would otherwise be infeasible by visual analysis because each slide can contain tens of thousands of cells. Automated detection also holds great promise for identifying objects the can be easily missed by visual examination, such as mitotic figures [149] or isolated tumour cells [19]. Segmentation of tissue structures, such as glands in colon tissue or ducts in breast tissue, enables exploration of morphological features that may be linked to cancer

grade and patient prognosis.

Cancer type and grade prediction: A routine task for the pathologist is to diagnose the type and grade of cancer because they are both major determinants of patient treatment [75, 44]. CPath can provide objective and reproducible measures- therefore helping to reduce diagnostic variability, as discussed in Section 1.2.5. For example, CPath can be used to automatically grade cancer, such as performing Gleason grading [14, 62], that may otherwise be subject to significant variation in pathologist diagnosis. Also, given a tissue that has been extracted from a specific organ, computational algorithms can automatically diagnose the cancer type, which is important because different types can be subject to different treatment regimens.

Prediction of prognosis: Diagnosing the cancer type and grade involves following a fixed set of guidelines. However, tasks such as the prediction of survival time, likelihood of recurrence and prediction of optimal treatment can be more complicated. Another advantage of using CPath algorithms is that they can automatically extract a representative set of features related to the task at hand. These features may subsequently be used to educate pathologists on the most diagnostic features for a given task. Furthermore, CPath algorithms can detect sub-visual features that may potentially enable overall superior diagnostic performance.

Challenges of Computational Pathology: As described above, computational methods are potentially advantageous for the analysis of digital histology slides, however there are various challenges that must be considered before developing such tools. First of all, as mentioned in Section 1.2.1, there can contain a large degree of variability in the appearance between different tissue slides. Therefore, we must ensure that algorithms are able to generalise well to new data, irrespective of their visual appearance. This is especially important if we expect an algorithm developed on a single cohort to perform well on another cohort with a slightly different tissue preparation procedure. Also, as mentioned in Section 1.3.1, WSIs are very large in size and therefore standard algorithms will not be able to work with the entire slide as input, due to computer memory constraints. As well as memory issues, WSIs typically take a long time to process and therefore developing efficient algorithms is a major challenge. This is an important consideration because it directly impacts the amount of diagnoses that a computational tool is able to provide in a given amount of time. Another challenge in CPath is for algorithms to accurately diagnose each tissue sample, given the complexity of histological patterns that may appear in any slide. For example, certain cells can be easily mistaken for others due to similarity in appearance and different cancer types may become difficult to

diagnose tumours become poorly differentiated. This challenge is highlighted by the difficulty for expert pathologists to reach consensus diagnosis for certain tasks.

1.4 Learning from Data

1.4.1 Machine Learning

Within CPath, Machine Learning (ML) is regularly used to solve some of the example tasks described in Section 1.3.2. ML is a branch of Artificial Intelligence (AI) that describes the process of learning from data to perform a task, rather than using a pre-determined equation. When using ML for image recognition, we define a set of N training images $\{\boldsymbol{x}^{(i)}\}_{i=1}^{N}$ and a function $f(\boldsymbol{x})$, that maps the input image to an output. Broadly speaking, if we provide target values $\{\boldsymbol{y}^{(i)}\}_{i=1}^{N}$, then we refer to the task as supervised learning; otherwise it is classed as unsupervised learning. Reinforcement learning is another type of ML that involves the process of determining appropriate actions to maximise a reward, but is not widely used for image recognition. In the supervised setting, we aim to learn a function such that for each example i the error (or loss) between $f(\boldsymbol{x}_i)$ and $\boldsymbol{y}_i$ is small. Specifically, when working with parametric ML models, we learn a set of parameters $\boldsymbol{W}$ to minimise:

$$\underset{\boldsymbol{W}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \ell(f(\boldsymbol{x}^{(i)}; \boldsymbol{W}), \boldsymbol{y}^{(i)}), \tag{1.1}$$

where ℓ is a pre-defined task-dependent loss function, such as cross entropy for a discrete target or mean squared error for a continuous target. After the learning (or training) process, the goal of a supervised ML model is to generate accurate predictions with a set of M unseen test images $\{\tilde{\boldsymbol{x}}^{(i)}\}_{i=1}^{M}$. In the unsupervised setting, we are not provided with target values and therefore the goal of an ML model may be to discover groups of similar examples in the data, determine the data distribution or reduce the data dimensionality [22].

1.4.2 Neural Networks

In this thesis, we mainly focus on the development of a subgroup of ML models, namely neural networks, in a supervised learning setting. The first mathematical model of an artificial neuron was developed back in 1943 by McCullough and Pitts. This was an extremely simple representation of a neuron, where a set of binary inputs were aggregated to give a binary response. In 1958, Rosenblatt developed the Perceptron to overcome some of the issues of the McCulloh and Pitts neuron. The

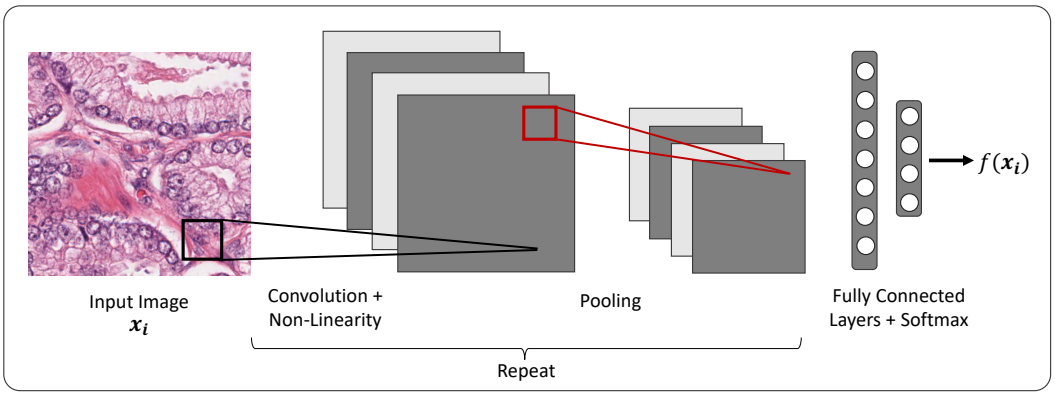


Figure 1.7: Convolutional neural network for classification. For convolution, the inner product is computed between the filter and input to get the pixel in the output feature map. This is shown by the black window. For pooling, a single statistic (such as the maximum or average) is computed in the red window to obtain the pixel in the output feature map. These operations are repeated for feature extraction and then fully connected layers are applied to give the final output.

Perceptron allowed a non-binary input and introduced the weight learning paradigm, that is now a central concept in modern neural networks. In 1986 the backpropagation algorithm [123] was developed and enabled neural networks with multiple layers to be effectively trained. As a result, neural networks could learn non-linear functions and in fact, were capable of learning any function due to the universal approximation theorem. These artificial neural networks (ANNs) are typically fully connected, where each input neuron is connected to every neuron in the next layer. As a result of this full connectivity, ANNs utilise a large number of parameters especially for high-dimensional input data, such as images, and are therefore often prone to overfitting.

1.4.3 Convolutional Neural Networks

The Neocognitron [49] was introduced in 1979 by Fukushima and employs a hierarchical, multi-layered design with the concept of local feature integration. This idea of feature *locality* was the source of inspiration for Convolutional Neural Networks (CNNs). In 1994, LeCun combined the idea of locality with backpropagation and developed a network that has become the backbone for many of today's AI algorithms [95]. LeCun recognised that images have translation symmetry and therefore chose to organise weights as 2D filters that are re-used over all spatial locations of the image. Then a convolution operation is performed, where the inner product is computed between the filter and image at each spatial position. This design results

in *translation-equivariance*, which means that a shift in the input leads to a proportional shift in the filter response, and significantly reduces the number of model parameters compared to ANNs.

Most modern CNNs consist of a series of convolution, pooling and non-linear operations that are applied sequentially to enable feature extraction. The output of a convolution between an input image and a filter is a 2D feature map, where its output spatial dimensions depend on whether padding was applied to the input image before convolving. Typically, a non-linear function is applied after each convolution, which enables complex functions to be learned by the network. The rectified linear unit (ReLU) is one of the most widely used functions in modern architectures, partly due to faster training times [86], that sets all negative outputs to 0. Pooling is often used to reduce the spatial dimensions of feature maps, which consequently increases a filter's field of view. This field of view is more commonly referred to as the *receptive field*. A pooling operation considers a small window of the input and computes a single statistic from all corresponding pixels, such as the maximum or average. This operation is then repeated over the input, where the stride of the window controls the output dimensions of the feature map. We display a simple CNN for classification in Figure 1.7 with a single convolution and pooling layer, which is repeated to automatically extract representative features. Following feature extraction, fully connected layers are used to obtain the final output, which is followed by a Softmax function to convert the output to a probability. When localising regions in the input image, a prediction is made per pixel, rather than for the entire image. This will be described in detail in Chapters 3 and 4 of the thesis.

With the increase in computing power, CNNs have since been developed to run on the GPU, helping overcome the issues of long processing times and has enabled the development of CNNs with many layers. For example, in 2012, Krizhevsky proposed a *deep* CNN [86] that was capable of excellent image recognition performance and has since inspired the development of a plethora of CNNs for computer vision. Now, CNNs are capable of achieving super-human performance in certain image recognition tasks [66, 71], motivating their usage in a wide range of modern applications.

1.4.4 CNNs in Computational Pathology

One area where CNNs have demonstrated recent success is the field of computational pathology. For example, they have shown their capability of reaching a greater diagnostic accuracy than the pathologist for breast cancer metastasis detection [19] and have achieved the best performance in multiple CPath image recognition contests

[149, 87, 135]. One contributing factor to the success of CNNs is their ability to exploit translation symmetry by reusing the convolution filter at all spatial positions of the input. However, image regions from WSIs are also symmetric under rotation because they can appear at any orientation with equal probability. Therefore it is desirable to extend the design of the CNN such that it is additionally equivariant to rotation to enable better feature map interpretability and potentially improve performance. In future, we believe that rotation-equivariant CNNs will become the standard choice for histopathology image analysis where rotational symmetry exists on a global scale.

CNNs take a single image as input, but as mentioned in Sections 1.3.1 and 1.3.2, WSIs are very large and therefore using the entire WSI at a high resolution along with the network parameters is often infeasible. To overcome this challenge, usually a *divide and conquer* strategy is employed in CPath. Specifically, the slides are first split into small image regions (or patches) for training the CNN. After training, unseen WSIs are then similarly divided into patches and a prediction is made for each patch by the trained CNN. If performing localisation/segmentation, then a prediction is made per pixel; otherwise a single prediction, such as the cancer grade or cancer type, is made. Then, patch-level predictions are aggregated to form a probability map, where a series of statistical measurements are typically calculated to obtain the overall slide-level prediction. In CPath, aggregated patch-level predictions can assist with the precise localisation of tissue components, such as nuclei and glands, enabling morphological features to be extracted. These features can be studied to better understand their link with patient outcome, which can help find cost-effective biomarkers and improve patient treatment.

1.5 Aims and Objectives

This thesis aims to develop automatic tools for the analysis of large-scale whole-slide images, that may help improve diagnostic pipelines in computational pathology. We initially develop a patch aggregation pipeline for WSI cancer type prediction to demonstrate the challenge of dealing with multi-gigapixel digitised tissue samples. The remainder of the thesis focuses on the investigation of techniques for accurate localisation of structures within the tissue, such as glands and nuclei, and the development of methods that exploit rotational symmetry within histology images. We mainly utilise algorithms in the area of machine learning and key concepts from group representation theory.

1.6 Main Contributions

- We introduce the challenge of working with multi-gigapixel WSIs and propose a patch aggregation approach for classifying non-small cell lung cancer WSIs into either lung adenocarcinoma or lung squamous cell carcinoma.
- We present the first network for simultaneous segmentation and classification of nuclei in histology images, named HoVer-Net. The network uses the concept of horizontal and vertical distance maps to separate clustered nuclei and uses a devoted upsampling branch for classification.
- We propose MILD-Net, a network for gland instance segmentation that counters the loss of information caused by max-pooling. In addition, the network uses atrous spatial pyramid pooling to segment glands with varying size and uses an uncertainty mechanism to highlight areas of ambiguity.
- We propose Rota-Net, which is a CNN for simultaneous segmentation of glands and lumen in colon histology images. Our proposed approach uses group convolutions to ensure that the network is equivariant to rotations of multiples of 90°.
- We propose Dense Steerable Filter CNNs (DSF-CNNs) that use group convolutions with multiple rotated copies of each filter in a densely connected framework. Each filter is defined as a linear combination of steerable basis filters, enabling exact rotation by any angle and decreasing the number of parameters compared to standard filters.

1.7 Thesis Organisation

Chapter 2: Patch Aggregation Computational Pathology. Cancer diagnosis is conventionally performed by visual examination of tissue sections under the microscope by a pathologist. In this chapter, we conduct a preliminary study that overcomes the difficulty of WSI cancer type classification by using a two-part *patch aggregation* strategy. First, we implement a deep learning (DL) model to classify input patches into different categories. Next, we extract a collection of statistical and morphological measurements from the labelled WSI and use a random forest regression model to classify the overall cancer type of each WSI. We apply our framework to the task of non-small cell lung cancer (NSCLC) classification and classify each WSI as either lung adenocarcinoma (LUAD) or lung squamous cell carcinoma (LUSC), which account for around 40% and 25-30% of all lung cancers

respectively. This task was part of the Computational Precision Medicine challenge at the MICCAI 2017 conference, where we achieved the highest classification accuracy with a score of 0.81. Our framework is not limited to cancer type prediction, but can also be used for other WSI classification tasks, such as cancer grading.

Chapter 3: HoVer-Net for Simultaneous Segmentation and Classificaion of Nuclei. The development of automated methods for nuclear segmentation and classification enables the quantitative analysis of tens of thousands of nuclei within a whole-slide pathology image, opening up possibilities of further analysis of large-scale nuclear morphometry. However, automated nuclear segmentation and classification is faced with a major challenge in that there are several different types of nuclei, some of them exhibiting large intra-class variability such as the tumour cells. Additionally, some of the nuclei are often clustered together. To address these challenges, we present a novel convolutional neural network for simultaneous nuclear segmentation and classification that leverages the instance-rich information encoded within the vertical and horizontal distances of nuclear pixels to their centres of mass. These distances are then utilised to separate clustered nuclei, resulting in an accurate segmentation, particularly in areas with overlapping instances. Then for each segmented instance, the network predicts the type of nucleus via a devoted upsampling branch. We demonstrate state-of-the-art performance compared to other methods on multiple independent multi-tissue histology image datasets.

Chapter 4: MILD-Net for Gland Instance Segmentation. The analysis of glandular morphology within colon histology images is an important step in determining the grade of colon cancer. Automated gland segmentation enables subsequent morphological analysis, yet remains a challenge due to variability in glandular appearance. To address this, we propose a fully convolutional neural network that counters the loss of information caused by max-pooling by re-introducing the original image at multiple points within the network and use atrous spatial pyramid pooling for multi-scale aggregation. To incorporate uncertainty, we introduce random transformations during test time for an enhanced segmentation result that simultaneously generates an uncertainty map, highlighting areas of ambiguity. We show that this map can be used to define a metric for disregarding predictions with high uncertainty. The proposed network achieves state-of-the-art performance on the GlaS challenge dataset and on a second independent colorectal adenocarcinoma dataset. In addition, we perform gland instance segmentation on whole-slide images from two further datasets to highlight the generalisabil-

ity of our method. As an extension, we introduce MILD-Net$^+$ for simultaneous gland and lumen segmentation, to increase the diagnostic power of the network.

Chapter 5: Exploiting Rotational Symmetry in Histology Images. Histology images are inherently symmetric under rotation, where each orientation is equally as likely to appear. However, this rotational symmetry is not widely utilised as prior knowledge in modern Convolutional Neural Networks (CNNs), resulting in *data hungry* models that learn independent features at each orientation. Allowing CNNs to be rotation-equivariant removes the necessity to learn this set of transformations from the data and frees up model capacity, allowing more discriminative features to be learned. In this chapter we explore the concept of rotation equivariance in CNNs for histology image analysis. First, we propose Rota-Net for simultaneous gland and lumen segmentation. This approach incorporates rotational symmetry into an encoder-decoder based network by utilising group equivariant convolutions with 90 degree filter rotations. Then, we propose Dense Steerable Filter CNNs (DSF-CNNs) that use group convolutions in a densely connected network, where each filter is defined as a linear combination of steerable basis filters. Utilising steerable filters enables rotation without artefacts and decreases the number of trainable parameters compared to standard filters. We observe that incorporating rotational symmetry leads to a boost in performance across multiple histology image datasets.

Chapter 6: Conclusions and Future Directions. This chapter summarises the main findings of the thesis and discusses future directions on how this work may be extended.

Chapter 2

Patch Aggregation Computational Pathology

Cancer diagnosis involves the visual examination of tissue morphology and cellular appearance under the microscope. In recent years, there has been a growing trend towards a digitised pathology workflow, where digital images are acquired from glass histology slides using a high-resolution scanning device and are now used in routine diagnosis [137]. The advent of digital pathology has led to a rise in computational pathology (CPath), where in particular machine learning (ML) and deep learning (DL) algorithms have shown great promise in assisting pathologists in diagnostic decision making. Whole-slide images (WSIs) obtained from scanning the original glass slides can be leveraged to develop algorithms for classification tasks, where a single label is assigned to each slide. However, this analysis is non-trivial due to the huge size of WSIs, where they can typically contain around 20 billion pixels. Therefore, standard ML and DL methods are unable to use the entire WSI as input, due to computer memory limitations. To overcome the difficulty of working with WSIs, the following classification pipeline is often used: (i) divide the WSI into smaller image patches, (ii) make predictions independently on each image patch and then (iii) predict the overall WSI label based on the aggregation of patch-level results. This *divide and conquer* strategy can be applied to many classification tasks within CPath including cancer type prediction, cancer grading and even prediction of genetic mutation [76].

In this chapter we present a framework for WSI classification to introduce the concept of patch aggregation and to display its potential in achieving a good performance for automated cancer diagnosis. Specifically, we propose a two-part approach to classify non-small cell lung cancer (NSCLC) WSIs into either lung

adenocarcinoma (LUAD) or lung squamous cell carcinoma (LUSC). First, we classify all input patches from an unseen WSI as either LUAD, LUSC or non-diagnostic (ND) using a deep neural network and obtain the corresponding probability maps for each class. Next, we extract a collection of statistical and morphological features from the LUAD and LUSC probability maps for input into a random forest regression model to classify each WSI. To the best of our knowledge, at the time of publication this was the first 3-class network for NSCLC classification that aims to classify each WSI into diagnostic and non-diagnostic areas. This task has been organised as part of the Computational Precision Medicine challenge at the MICCAI 2017 conference, where we achieved the greatest accuracy with a score of 0.81. Our pipeline is not limited to the task of NSCLC classification, but can be used to diagnose between other cancer types and for cancer grading.

2.1 Non-Small Cell Lung Cancer Classification

Distinguishing between LUAD and LUSC is an important task because it can help determine patient treatment [141]. LUSC accounts for about 25-30% of all lung cancers and originates in the tissue that lines the air passages within the lung. In well differentiated LUSC, histological features that can be observed include inter-cellular bridging and keratinisation in pearl formation. LUAD originates in the mucus-secreting glandular cells within the lung and accounts for about 40% of all lung cancers. LUAD displays large variability in its appearance, where a given tissue section can be of the following major sub-types: acinar, papillary, micro-papillary, lepidic and solid tumour growth patterns. It is interesting to note that over 80% of LUAD cases today are diagnosed as a mixed sub-type, consisting of two or more histological sub-types. Some typical examples from both LUSC and LUAD can be seen in parts (a) and (b) of Figure 2.1. We also display some non-tumour regions that contain alveoli, connective tissue, immune cells and fat. Despite the importance of distinguishing between NSCLC histological types, the task is non-trivial for poorly differentiated cases where typical morphological features are infrequent. Furthermore, manual inspection and analysis of whole-slide image (WSIs) to detect these types of lung cancer is a labor-intensive, subjective and time-consuming task particularly when the workload is high.

Similar to traditional examination, the automation of this task remains a challenge because typical histological features are not as obvious in poorly differentiated tumours and there is a high level of intra-class heterogeneity. Figure 2.1 highlights the difficulty in distinguishing between lung adenocarcinoma and lung

squamous cell carcinoma when diagnostic features are uncommon. For example, the top two LUAD image regions from part (a) of Figure 2.1 display distinct growth patterns, whereas the bottom two regions could easily be confused with LUSC.

2.2 Related Work

WSIs can be up to 150,000×100,000 pixels in size at their highest magnification level, which provides a key challenge for most ML and DL models because the entire slide will typically not fit into memory. Therefore, usually a patch aggregation strategy is utilised to yield the overall WSI prediction. This principle has been used in a variety of tasks in CPath, including breast cancer metastasis detection [19, 98], Gleason grading [14] and colorectal cancer grading [129].

There have been a number of recent methods for automated NSCLC classification. For example, Yu *et al.* [166] extracted a range of quantitative image features from tissue regions and used a series of classical ML techniques to classify each WSI. Although hand crafted approaches perform well, there is a growing trend towards deep learning approaches, where networks are capable of learning a strong feature representation. In particular, deep convolutional neural networks (CNNs) [66, 140, 133, 71] exploit translation symmetry within images by using a weight-sharing strategy in the model architecture, leading to remarkable performance in image recognition tasks [41]. Coudray *et al.* [37] utilised the above mentioned patch-based approach for NSCLC classification using deep learning, but in addition predicted the ten most commonly mutated genes. For lung cancer classification, the authors used an Inception v3 [140] network architecture to classify input patches into LUAD, LUSC and normal. The authors assumed that all patches within each WSI had the same label and therefore did not differentiate between diagnostic and non-diagnostic regions. This method may result in a large amount false positives in non-diagnostic regions and training may take a long time to converge. Hou *et al.* [69] trained a patch-level classifier to classify glioma and NSCLC WSIs into different cancer types. This was done by aggregating discriminative patch-level predictions from a deep network using either a multi-class logistic regression model or support vector machine. The selection of discriminative patches was done in a weakly supervised manner, where an expectation-maximisation approach was used to iteratively select patches. These patches were then fed into a conventional two-class CNN to classify input patches as LUAD or LUSC. The authors counter the problem of differentiating between diagnostic and non-diagnostic regions by only considering discriminative patches. Although successful, this technique would likely

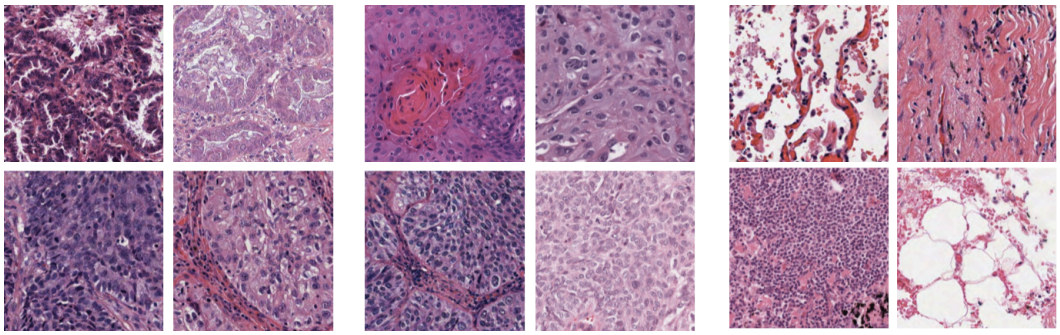


Figure 2.1: Examples of typical patches from each class. All patches are displayed at 256×256 at 10× resolution. (a): LUAD patches; (b) LUSC patches; (c) ND patches.

fail if presented with a small unrepresentative dataset.

2.3 Methods

In this section, we present our proposed patch aggregation approach that we apply to the task of NSCLC classification. The section is broken down into three main parts: (i) dataset description; (ii) deep learning framework for patch based classification; (iii) random forest regression model for classifying a WSI as LUAD or LUSC. An overview of the WSI classification framework can be viewed in Figure 2.2.

2.3.1 The Dataset

As part of the Computational Precision Medicine (CPM) challenge [151] at the MICCAI 2017 conference, we used a total of 64 Hematoxylin and Eosin (H&E) NSCLC WSIs that were split into 32 training and 32 test images. Ground truth was supplied for the training images that gave the cancer type of each WSI, whereas this ground truth was held back by the challenge organisers for the test images. We had an even breakdown of NSCLC images in both the training and the test set, giving a total of 32 LUAD slides and 32 LUSC slides. We divided our dataset so that we had 24 WSIs for training and 8 for validation, with 4 validation images taken from LUAD and LUSC respectively. We extracted a 3 class dataset consisting of patches of size 256×256 at 20× magnification, from non-exhaustive labelled regions, confirmed by an expert pathologist (AK). This 3 class dataset consisted of LUAD, LUSC and non-diagnostic areas (ND). We considered regions containing tumour and tumour associated stroma to be diagnostic. Here, we consider growth patterns

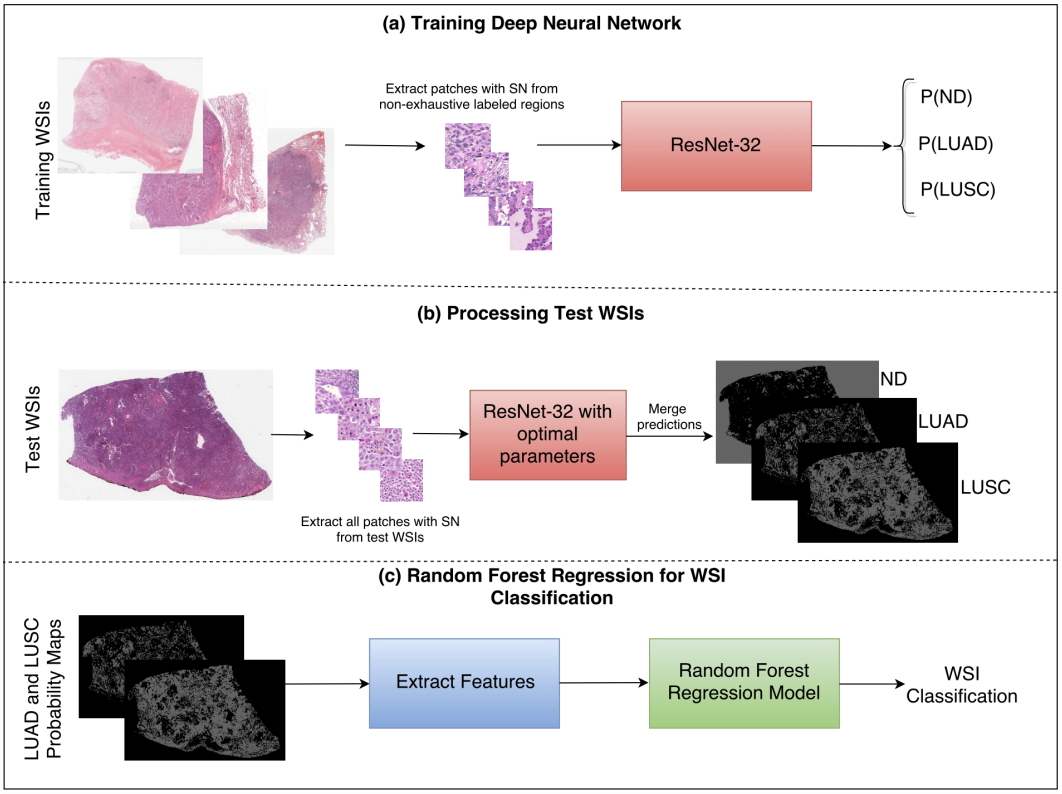


Figure 2.2: Overview of the patch aggregation approach for NSCLC WSI classification. (a) Training a patch-based classifier (b) Processing unseen WSIs. (c) Random forest regression model for WSI classification. SN stands for stain normalisation via method of Reinhard [120].

and keratin pearls to be tumour. Non-diagnostic regions included fat, lymphocytes, blood vessels, alveoli, red blood cells, normal stroma, cartilage and necrosis. We considered necrosis to be non-diagnostic because, despite LUSC generally having more necrotic areas than LUAD, it is not indicative of lung squamous cell carcinoma on a patch-by-patch basis. Overall, our network is optimised on 65,788 training image patches.

Despite all slides being stained with H&E, there was a high level of stain variability from image to image. Therefore, we applied Reinhard [120] stain normalisation to all images to limit the reduce the variation in the stain appearance. During training we performed random crop, flip and rotation data augmentation to make the network invariant to these transformations. After performing a random crop to all input patches, we were left with a patch size of 224×224.

2.3.2 Deep Neural Network for Patch-Based Classification

Convolutional Neural Network

An increase in the amount of labelled data coupled with a surge in computing power has allowed deep CNNs to achieve state-of-the-art performance in computer vision tasks. The hierarchical architecture of such networks allow them to have a strong representational power, where the complexity of learned features increases with the depth of the network. The proposed network f is a composition of a sequence of L functions of layers $(f_1, ..., f_L)$ that maps an input vector $\boldsymbol{x}$ to an output vector $\boldsymbol{y}$, i.e,

$$\boldsymbol{y} = f(\boldsymbol{x}; \boldsymbol{w}_1, ...\boldsymbol{w}_L) = f_L(\cdot; \boldsymbol{w}_L) \circ f_{L-1}(\cdot; \boldsymbol{w}_{L-1}) \circ ...f_2(\cdot; \boldsymbol{w}_2) \circ f_1(\boldsymbol{x}; \boldsymbol{w}_1) \quad (2.1)$$

where $\mathbf{w}_l$ is the weight and bias vector for the l^{th} layer f_l. In practice, f_l most commonly performs one of the following operations: a) convolution with a set of filters; b) spatial pooling; and c) non-linear activation.

Given a set of training data $\{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{N}$, we can estimate the vectors $\mathbf{w}_1, .., \mathbf{w}_L$ by solving:

$$\underset{\boldsymbol{w}_1 ... \boldsymbol{w}_L}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \ell(f(\boldsymbol{x}^{(i)}; \boldsymbol{w}_1, ...\boldsymbol{w}_L), \boldsymbol{y}^{(i)}), \quad (2.2)$$

where l is the defined loss function. We perform numerical optimisation of (2) conventionally via the back-propagation algorithm and stochastic gradient descent methods.

In addition to the above operations, residual networks (ResNets) [66] have recently been proposed that enable networks to be trained deeper and as a result, benefit from a greater accuracy. Current-state-of-the-art networks[66, 140, 133, 71] indicate that network depth is of crucial importance, yet within conventional CNNs, accuracy gets saturated and then degrades rapidly as the depth becomes significantly large. The intuition behind a residual network is that it is easier to optimise the residual mapping than to optimize the original unreferenced mapping. Residual units are the core components ResNets and contain feed-forward skip connections that perform identity mapping without adding any extra parameters. These connections propagate the gradient throughout the model, which in turn enables the network to be trained deeper, often achieving greater accuracy. An example of a residual unit can be seen in part (b) of Figure 2.3.

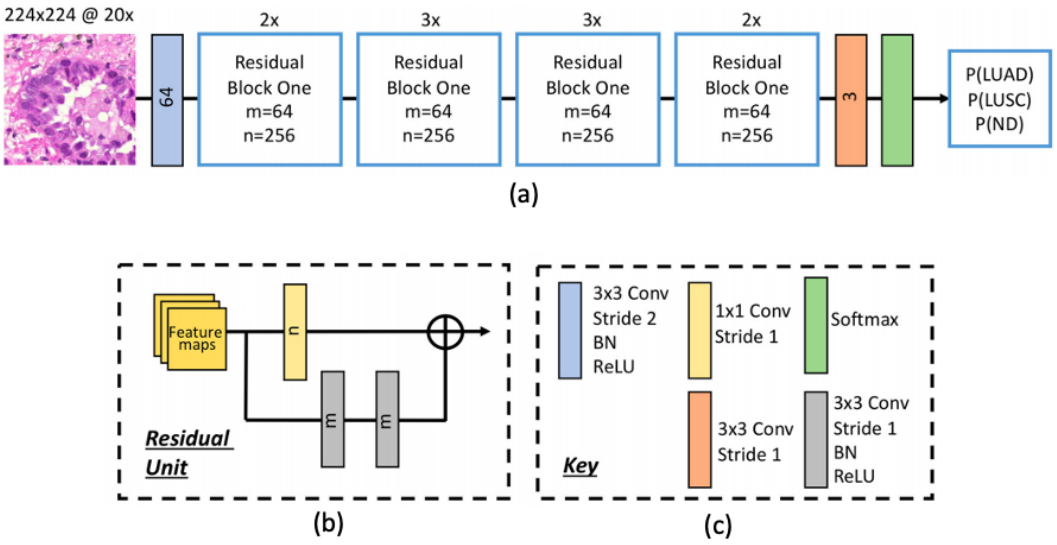


Figure 2.3: The proposed deep convolutional neural network. (a) Network architecture, (b) residual block, (c) key. Within the residual block, $\oplus$ refers to a summation operator.

Proposed network architecture

Inspired by the success of ResNet [66] in image-recognition tasks [41], we implemented a deep neural network with residual blocks at its core to classify NSCLC input patches. This network architecture is a variant of ResNet50, as described by He *et al.* [66], but we use a 3×3 kernel as opposed to a 7×7 kernel during the first convolution to reduce the number of parameters and then further reduce the number of parameters in the remainder of the network. Reducing the amount of parameters helps the network to generalise better to new data and reduces the risk of overfitting. In order to reduce the amount of parameters, we modified ResNet50 [66] by reducing the amount of residual blocks throughout the network so that we had 32 layers as opposed to 50. Due to the high variability between images, and therefore between the training and validation set, consideration for preventing overfitting is crucial. Figure 2.3 gives an overview of the network architecture.

Once training was complete, we selected the optimal epoch corresponding to the greatest average validation accuracy and processed patches from each test WSI. This resulted in three probability maps; one for each class.

2.3.3 Extraction of Statistical and Morphological Features

For classifying each WSI as either LUAD or LUSC, we extracted features from both respective probability maps. We explored two post processing techniques: max voting and a random forest regression model. Max voting simply assigns the class of the WSI to be class with the largest amount of positive patches in its corresponding probability map. Therefore, max voting only requires the positive patch count for both the LUAD and LUSC probability maps in order to make a classification. For the random forest regression model, we extracted 50 statistical and morphological features from both the LUAD and LUSC training probability maps and then selected the top 25 features based on class separability. We gained the training probability maps by processing each training WSI with a late epoch. This ensured that the network had overfit to the training data and gave a good segmentation of LUAD and LUSC diagnostic regions. In other words, using this method allowed us to transition from a non-exhaustive to an exhaustive labelled probability map. Once the model was trained with these features, they were then input as features into the random forest regression model. Statistical features that were extracted included: mean, median and variance of the probability maps. Morphological features that were extracted included the size of the top five connected components at different thresholds.

2.3.4 Random Forest Regression Model

An ensemble method is a collection of classifiers that are combined together to give improved results. An example of such an ensemble method is a random forest, where multiple decision trees are combined to yield a greater classification accuracy. Decision trees continuously split the input data, according to a certain parameter until a criterion is met. Specifically, a random forest regression model fits a number of decision trees on various sub-samples of the data and then calculates the mean output of all decision trees. We optimised our random forest model by selecting an ensemble of 10 bagged trees, randomly selecting one third of variables for each decision split and setting the minimum leaf size as 5. We finally selected a threshold value to convert the output of the random forest regression model into a binary value, indicating whether the WSI was LUAD or LUSC.

Table 2.1: Patch-Level accuracy. LUAD refers to lung adenocarcinoma, LUSC refers to lung squamous cell carcinoma, ND refers to non-diagnostic area.

Model	LUAD	LUSC	ND	Average
VGG-16 [133]	0.634	0.663	0.826	0.708
Inception-v3 [140]	0.623	0.733	0.924	0.760
ResNet50 [66]	0.601	0.597	0.889	0.695
ResNet32	0.702	0.849	0.742	0.764

2.4 Results

Table 2.1 summarises the experiments we carried out for classification of input patches into LUAD, LUSC and ND. We chose to train with VGG16 [133], Inception-v3 [140] and ResNet [66] because of their state-of-the-art performance in recent image recognition tasks [41]. During training, we could see that our networks were overfitting. This was because of two reasons: (i) The networks architectures that were used have been optimised for large-scale computer vision tasks with millions of images and thousands of classes and (ii) there is a large variability between the training set and the validation set. With such a small and visually diverse dataset, (ii) is hard to avoid and therefore we modify the network architecture to counter the problem of overfitting. Modification of ResNet50 to give ResNet32 helped alleviate the problem of overfitting and gave the best patch-level performance. Despite only achieving 0.4% greater accuracy than Inception-v3, ResNet32 resulted in a significantly greater average LUAD and LUSC patch-level accuracy. The average LUAD and LUSC patch-level accuracy for Inception-v3 was 0.678, whereas the average accuracy for ResNet32 was 0.776. As a result, we chose to use ResNet32 for processing images in the test set. Figure 2.4 and Figure 2.5 show two example test WSIs with their overlaid probability maps. Green regions show regions classified as LUSC, blue/purple regions show regions classified as LUAD and yellow/orange regions show regions classified as ND. Here, we observe that the predicted ND regions in the Figure 2.4 consist of normal stromal, blood vessel and alveoli regions as expected. Also, our algorithm recognises the growth patterns present and therefore predicts the tumour area as LUAD. Figure 2.5 primarily consists of LUSC tumour and therefore the prediction is relatively homogeneous.

Table 2.2 shows the overall accuracy for NSCLC WSI classification, as processed by the challenge organisers. We observe that using the random forest re-

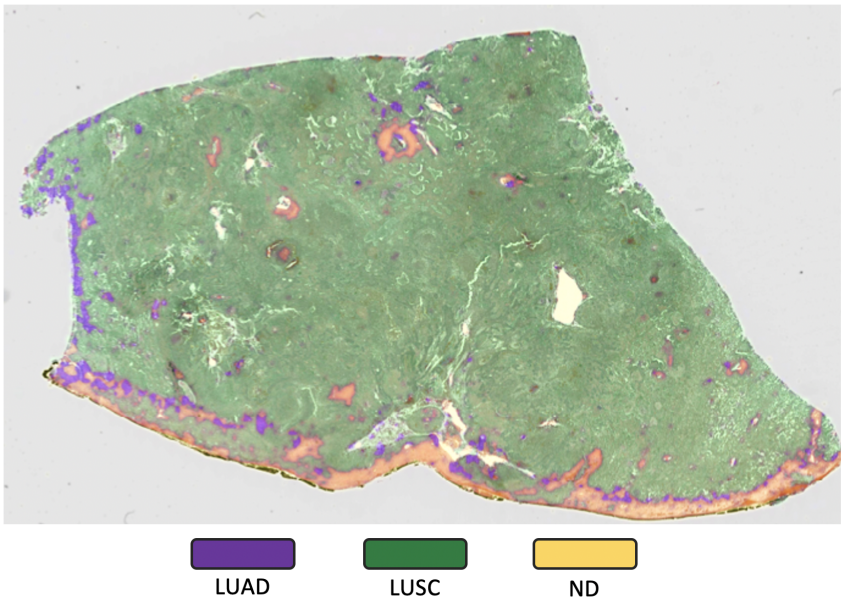


Figure 2.4: Unseen LUAD WSI with overlaid probability map. Blue/purple indicates predicted LUAD regions, green indicates predicted LSUC region and yellow/orange indicates a predicted ND region.

gression model with statistical and morphological features from the labelled WSI increases the classification accuracy. Max voting is sufficient when either LUAD or LUSC is a dominant class within the labelled WSI, but when there is no obvious dominant class, the random forest regression model increases performance. This is because the features used as input to the random forest model are tailored to the task of NSCLC classification and can therefore better differentiate between each cancer type. We may see a greater effect of the RF post-processing technique when applying it to a more challenging task, such as differentiating between the different histological sub-types in LUAD.

Table 2.2: Overall WSI classification accuracy using two different post-processing techniques. MV and RF refer to majority voting and random forest regression model respectively.

Method	**Accuracy**
ResNet32-MV	0.78
ResNet32-RF	0.81

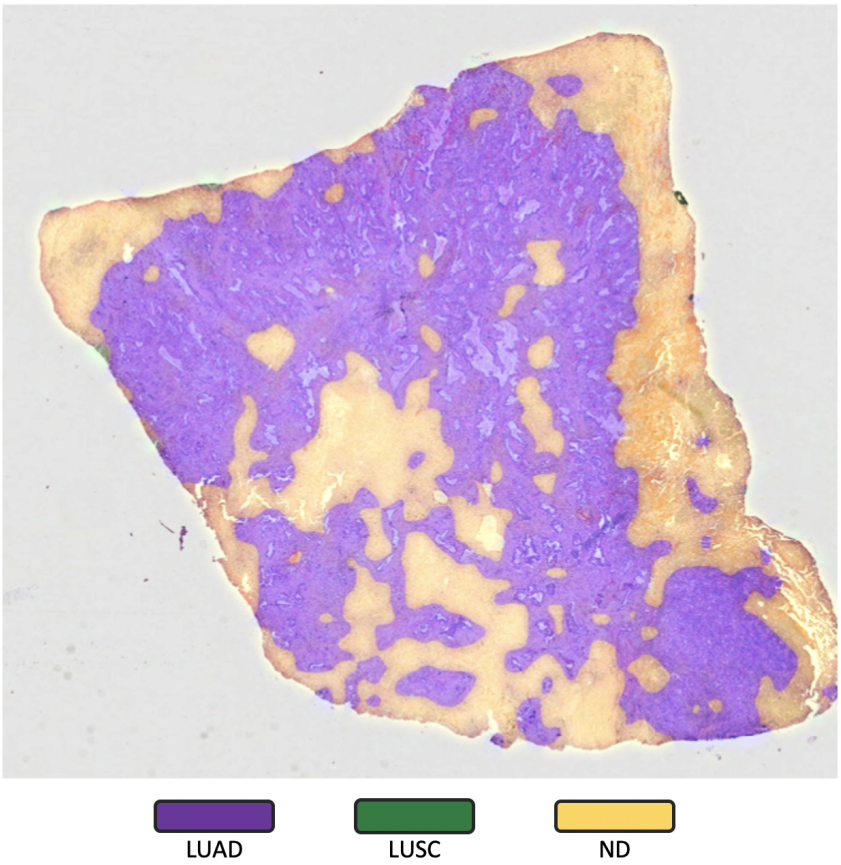


Figure 2.5: Unseen LUAD WSI with overlaid probability map. Blue/purple indicates predicted LUAD regions, green indicates predicted LSUC region and yellow/orange indicates a predicted ND region.

2.5 Discussion and Conclusions

This chapter presented a patch aggregation pipeline for large-scale WSI classification, where in particular we automatically classified NSCLC images as either LUAD or LUSC. In the first step of our classification framework, we implemented a deep neural network to classify input patches as LUAD, lung squamous cell carcinoma or non-diagnostic regions. In the second step, after processing each image, we extracted a collection of statistical and morphological features from the LUAD and LUSC probability maps. These features were then used as input into a random forest regression model to classify each WSI as lung adenocarcinoma or lung squamous cell carcinoma. Overall we achieved the greatest accuracy with a score of 0.81 as part of the Computational Precision Medicine challenge at MICCAI 2017. Especially given the limitation of the dataset, classifying NSCLC WSIs into diagnostic

and non-diagnostic regions seems to be of importance. Our proposed pipeline is not limited to NSCLC classification and can be used to diagnose a range of different cancer types and grades.

The consideration of contextual information can provide additional assistance in classification tasks within computational pathology [20, 8]. For example, growth patterns in LUAD cases and how the tumour grows with the stroma is of great importance when classifying NSCLC cases. These patterns are often very hard to visualize in a 224×224 patch at 20× resolution. In future work, developing our proposed network to include more contextual information may improve patch-level accuracy and therefore overall classification accuracy. Also, our network was trained on a relatively small number of WSIs. Training on a much larger number of WSIs will enable the deep learning model to better discriminate between challenging image regions and will therefore give better performance.

Our framework relies on pathologist annotation of LUAD, LUSC and ND regions. However, as we use more and more WSIs within our framework, the annotation burden will increase dramatically and will inevitably become impractical. Therefore, one area of interest is *weakly supervised learning* that instead only requires a weak annotation often in the form of a single slide-level label for CPath. For example, only the NSCLC diagnosis is required and therefore a vast amount of data can be used. This can be modelled as a multiple-instance learning problem, where each WSI can be labelled as a bag containing its corresponding patches (or instances for multiple-instance learning) and the goal is to predict the label for unseen bags. Previous work [23] has shown that multiple-instance learning algorithms enable CPath to be done at scale and allows terabytes of data to be used without the need for time-consuming pixel-level annotation.

Chapter 3

HoVer-Network for Nuclear Instance Segmentation

Current manual assessment of Haematoxylin and Eosin (H&E) stained histology slides suffers from low throughput and is naturally prone to intra- and inter-observer variability [43]. To overcome the difficulty in visual assessment of tissue slides, there is a growing interest in digital pathology (DPath), where digitised whole-slide images (WSIs) are acquired from glass histology slides using a scanning device. This permits efficient processing, analysis and management of the tissue specimens [109]. Each WSI contains tens of thousands of nuclei of various types, which can be further analysed in a systematic manner and used for predicting clinical outcome. Here, the type of nucleus refers to the cell type in which it is located. For example, nuclear features can be used to predict survival [12] and also for diagnosing the grade and type of disease [104]. Also, efficient and accurate detection and segmentation of nuclei can facilitate good quality tissue segmentation [136, 74], which can in turn not only facilitate the quantification of WSIs but may also serve as an important step in understanding how each tissue component contributes to disease. In order to use nuclear features for downstream analysis within computational pathology, nuclear segmentation must be carried out as an initial step. However, this remains a challenge because nuclei display a high level of heterogeneity and there is significant inter- and intra-instance variability in the shape, size and chromatin pattern between and within different cell types, disease types or even from one region to another within a single tissue sample. Tumour nuclei, in particular, tend to be present in clusters, which gives rise to many overlapping instances, providing a further challenge for automated segmentation, due to the difficulty of separating neighbouring instances. This separation is not just important for accurate feature extraction,

but also for cell counting. For example, tumour budding (TB) is defined as discrete clusters of up to four cancer cells [107] and without accurate nuclear instance segmentation, automatic recognition of TB is difficult.

As well as extracting each individual nucleus, determining the type of each nucleus can increase the diagnostic potential of current DPath pipelines. For example, accurately classifying each nucleus to be from tumour or lymphocyte enables downstream analysis of tumour infiltrating lymphocytes (TILs), which have been shown to be predictive of cancer recurrence [36]. Yet, similar to nuclear segmentation, classifying the type of each nucleus is difficult, due to the high variance of nuclear appearance within each WSI. Typically, nuclei are classified using two disjoint models: one for detecting each nucleus and then another for performing nuclear classification [130, 153]. However, it would be preferable to utilise a single unified model for nuclear instance segmentation and classification.

In this chapter, we present a deep learning approach[1] for simultaneous segmentation and classification of nuclear instances in histology images. The network is based on the prediction of horizontal and vertical distances (and hence the name HoVer-Net) of nuclear pixels to their centres of mass, which are subsequently leveraged to separate clustered nuclei. For each segmented instance, the nuclear type is subsequently determined via a dedicated upsampling branch. To the best of our knowledge, this is the first approach that achieves instance segmentation and classification within the same network. We present comparative results on six independent multi-tissue histology image datasets and demonstrate state-of-the-art performance compared to other recently proposed methods. The main contributions of this work are listed as follows:

- A novel network, targeted at simultaneous segmentation and classification of nuclei, where horizontal and vertical distance map predictions separate clustered nuclei.
- We show that the proposed HoVer-Net achieves state-of-the-art performance on multiple H&E histology image datasets, as compared to over a dozen recently published methods.
- An interpretable and reliable evaluation framework that effectively quantifies nuclear segmentation performance and overcomes the limitations of existing performance measures.

[1]Model code available at: `https://github.com/vqdang/hover\_net`

- A new dataset[2] of 24,319 exhaustively annotated nuclei within 41 colorectal adenocarcinoma image tiles.

3.1 Related Work

3.1.1 Nuclear Instance Segmentation

Within the current literature, **energy-based** methods, in particular the watershed algorithm, have been widely utilised to segment nuclear instances. For example, [164] used thresholding to obtain the markers and the energy landscape as input for watershed to extract the nuclear instances. Nonetheless, thresholding relies on a consistent difference in intensity between the nuclei and background, which does not hold for more complex images and hence often produces unreliable results. Various approaches have tried to provide an improved marker for marker-controlled watershed. [29] used active contours to obtain the markers. [148] used a series of morphological operations to generate the energy landscape. However, these methods rely on the predefined geometry of the nuclei to generate the markers, which determines the overall accuracy of each method. Notably, [11] avoided the trouble of refining the markers for watershed by designing a method that relies solely on the energy landscape. They combined an active contour approach with nuclear shape modelling via a level-set method to obtain the nuclear instances. Despite its widespread usage, obtaining sufficiently strong markers for watershed is a non-trivial task. Some methods have departed from the energy-based approach by utilising the geometry of the nuclei. For instance, [157], [93] and [89] computed the concavity of nuclear clusters, while [97] used eclipse-fitting to separate the clusters. However, this assumes a predefined shape, which does not encompass the natural diversity of the nuclei. In addition, these methods tend to be sensitive to the choice of manually selected parameters.

Recently, **deep learning** methods have received a surge of interest due to their superior performance in many computer vision tasks [101, 131, 94]. These approaches are capable of automatically extracting a representative set of features, that strongly correlate with the task at hand. As a result, they are preferable to hand-crafted approaches, that rely on a selection of pre-defined features. Inspired by the Fully Convolutional Network (FCN) [103], U-Net [121] has been successfully applied to numerous segmentation tasks in medical image analysis. The network has an encoder-decoder design with skip connections to incorporate low-level in-

[2]The CoNSeP dataset for nuclear segmentation is available at `https://warwick.ac.uk/fac/sci/dcs/research/tia/data/`.

formation and uses a **weighted loss function** to assist separation of instances. However, it often struggles to split neighbouring instances and is highly sensitive to pre-defined parameters in the weighted loss function. A more recently proposed method in Micro-Net [119] extends U-Net by utilising an enhanced network architecture with weighted loss. The network processes the input at multiple resolutions and as a result, gains robustness against nuclei with varying size. In [59], the authors developed a network that is robust to stain variations in H&E images by introducing a weighted loss function that is sensitive to the Haematoxylin intensity within the image.

Other methods exploit information about the nuclear **contour** (or boundary) within the network, such as DCAN [27] that utilised a dual architecture that outputs the nuclear cluster and the nuclear contour as two separate prediction maps. Instance segmentation is then achieved by subtracting the contour from the nuclear cluster prediction. [39] proposed a network to predict the inner nuclear instance, the nuclear contour and the background. The network utilised a customised weighted loss function based on the relative position of pixels within the image to improve and stabilise the inner nuclei and contour prediction. Some other methods have also utilised the nuclear contour to achieve instance segmentation. For example, [88] employed a deep learning technique for labelling the nuclei and the contours, followed by a region growing approach to extract the final instances. [82] used the contour predictions as input into a further network for segmentation refinement. [169] proposed CIA-Net, that utilises a multi-level information aggregation module between two task-specific decoders, where each decoder segments either the nuclei or the contours. A Deep Residual Aggregation Network (DRAN) was proposed by [151] that uses a multi-scale strategy, incorporating both the nuclei and nuclear contours to accurately segment nuclei.

There have been various other methods to achieve instance separation. Instead of considering the contour, [113] proposed a deep learning approach to detect superior markers for watershed by regressing the nuclear **distance map**. Therefore, the network avoids making a prediction for areas with indistinct contours.

In line with these developments, the field of instance segmentation within natural images is also rapidly progressing and have had a significant influence on nuclear instance segmentation methods. A notable example is Mask-RCNN [65], where instance segmentation approach is achieved by first predicting candidate regions likely to contain an object and then deep learning based segmentation within those proposed regions.

3.1.2 Nuclear Classification

As well as performing instance segmentation, it is desirable to determine the *type* of each nucleus to facilitate and improve downstream analysis. It is possible for current models to differentiate between certain nuclear types in H&E, however sub-typing of lymphocytes is an extremely hard task due to the high levels of similarity in morphological appearance between T and B lymphocytes. Typically, classifying each nucleus is done via a two-stage approach, where the first step involves either nuclear segmentation or nuclear detection. When segmentation is used as the initial step, a series of morphological and textural features are extracted from each instance, which are then used within a classifier to determine the nuclei classes. For example, [114] classified nuclei within H&E stained breast cancer images as either tumour, lymphocyte or stromal based on their morphological features. [167] performed nuclear segmentation and then classified each nucleus with AdaBoost classifier, utilising the intensity, morphology and texture of nuclei as features. Otherwise, detection is performed as an initial step and a patch centred at the point of detection is fed into a classifier, to predict the type of nucleus. [134] proposed a spatially constrained CNN, that initially detects all nuclei and then for each nucleus an ensemble of associated patches are fed into a CNN to predict the type to be either epithelial, inflammatory, fibroblast or miscellaneous.

3.2 Methods

Our overall framework for automatic nuclear instance segmentation and classification can be observed in Fig. 3.1 and the proposed network in Fig. 3.2. Here, nuclear pixels are first detected and then, a tailored post-processing pipeline is used to simultaneously segment nuclear instances and obtain the corresponding nuclear types. The framework is based upon the horizontal and vertical distance maps, which can be seen in Fig. 3.3. In the figure, each nuclear pixel denotes either the horizontal or vertical distance of pixels to their centres of mass.

3.2.1 Network Architecture

In order to extract a strong and representative set of features, we employ a deep neural network. The feature extraction component of the network is inspired by the pre-activated residual network with 50 layers [67] (Preact-ResNet50), due to its excellent performance in recent computer vision tasks [41] and robustness against input perturbation [13]. Compared to the standard Preact-ResNet50 implementation,

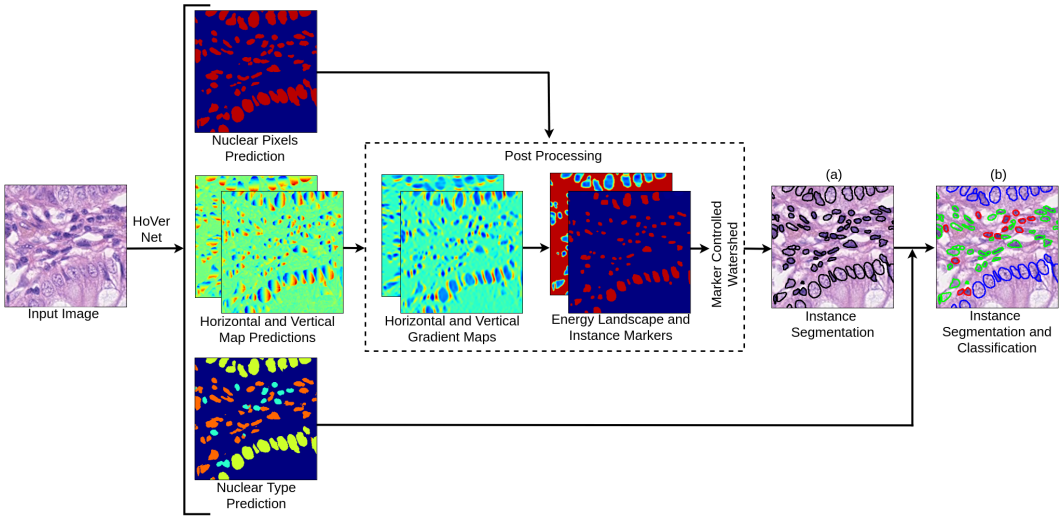


Figure 3.1: Overview of the proposed workflow for simultaneous nuclear instance segmentation and classification.

we reduce the total downsampling factor from 32 to 8 by using a stride of 1 in the first convolution and removing the subsequent max-pooling operation. This ensures that there is no immediate loss of information that is important for performing an accurate segmentation. Various residual units are applied throughout the network at different downsampling levels. A series of consecutive residual units is denoted as a residual block. The number of residual units within each residual block is 3, 4, 6 and 3 that are applied at downsampling levels 1, 2, 4 and 8 respectively. For clarity, a downsampling level of 2 means that the input has a reduction in the spatial resolution by a factor of 2.

Following Preact-ResNet50, we perform nearest neighbour upsampling via three distinct branches to simultaneously obtain accurate nuclear instance segmentation and classification. We name the corresponding branches: (i) nuclear pixel (NP) branch; (ii) HoVer branch and (iii) nuclear classification (NC) branch. The NP branch predicts whether or not a pixel belongs to the nuclei or background, whereas the HoVer branch predicts the horizontal and vertical distances of nuclear pixels to their centres of mass. Then, the NC branch predicts the type of nucleus for each pixel. In particular, the NP and HoVer branches jointly achieve nuclear instance segmentation by first separating nuclear pixels from the background (NP branch) and then separating touching nuclei (HoVer branch). The NC branch determines the type of each nucleus by aggregating the pixel-level nuclear type predictions within each instance.

All three upsampling branches utilise the same architectural design, which

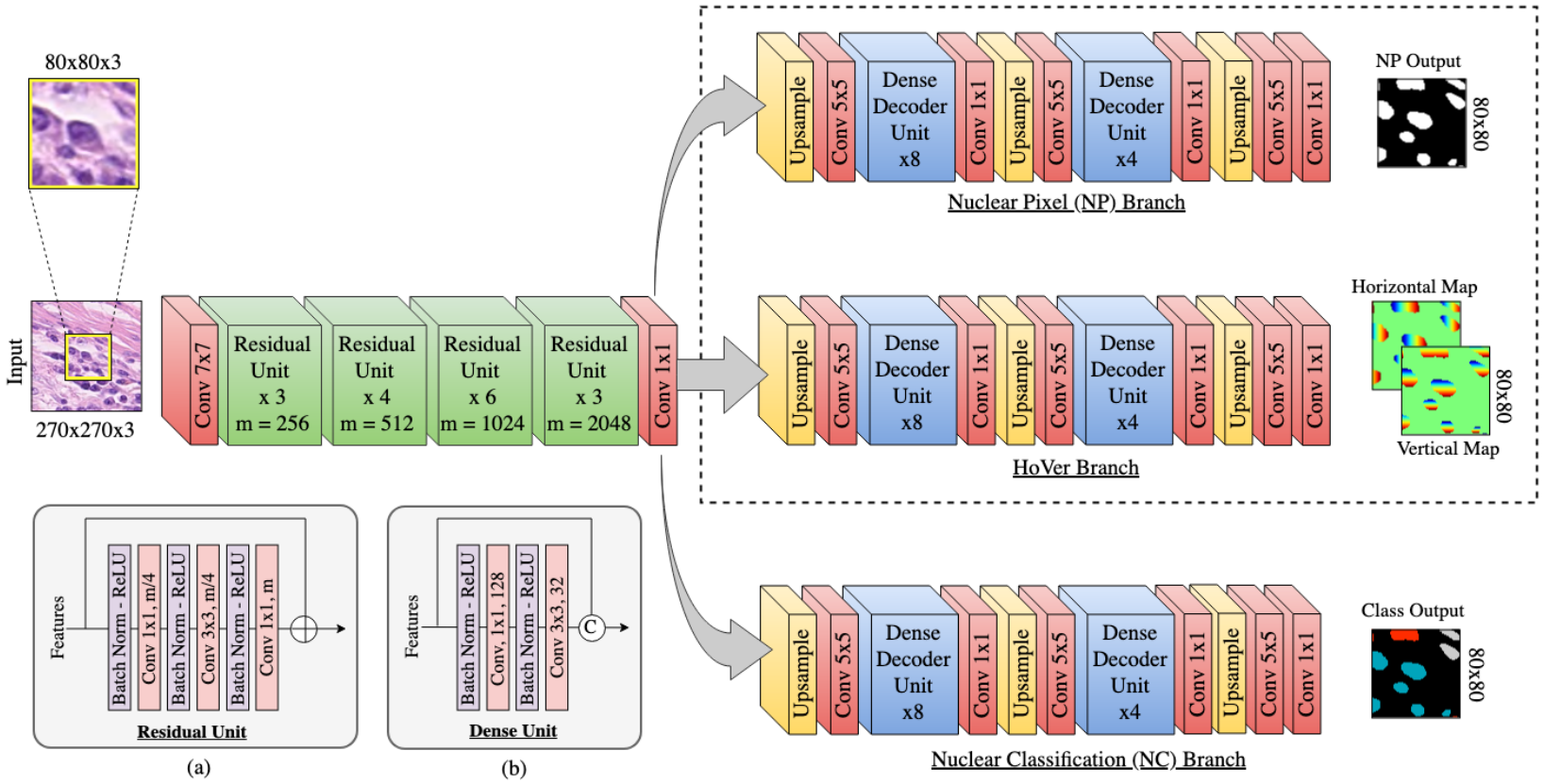


Figure 3.2: Overview of the proposed network architecture for simultaneous nuclear segmentation and classification.

consists of a series of upsampling operations and densely connected units [71] (or dense units). By stacking multiple and relatively cheap dense units, we build a large receptive field with minimal parameters, compared to using a single convolution with a larger kernel size and we ensure efficient gradient propagation. We use skip connections [121] to incorporate features from the encoder, but utilise summation as opposed to concatenation. The consideration of low-level information is particularly important in segmentation tasks, where we aim to precisely delineate the object boundaries. We use dense units after the first and second upsampling operations, where the number of units is 4 and 8 respectively. Valid convolution is performed throughout the two upsampling branches to prevent poor predictions at the boundary. This results in the size of the output being smaller than the size of the input. As opposed to using a dedicated network for each task, a shared encoder makes it possible to train the nuclear instance segmentation and classification model end-to-end and therefore, reduce the total training time. Furthermore, a shared encoder can also take advantage of the shared information across multiple tasks and thus, help to improve the model performance on all tasks.

Finally, if we do not have the classification labels of the nuclei, only the NP and HoVer upsampling branches are considered. Otherwise, we consider all three upsampling branches and perform simultaneous nuclear instance segmentation and classification.

We display an overview of the network architecture in Fig. 3.2, where the spatial dimension of the input is 270×270 and the output dimension of each branch is 80×80. The dashed box within Fig. 3.2 highlights the branches for nuclear instance segmentation. Additionally, we also show a residual unit and a dense unit within Fig. 3.2a and Fig. 3.2b. We denote m as the number of feature maps within each convolution of a given residual unit. At each down sampling level, from left to right, m=256, 512, 1024, 2048 respectively. We keep a fixed amount of feature maps within each dense unit throughout the two branches as shown in Fig. 3.2c.

Loss Function

The proposed network design has 4 different sets of weights: w_0, w_1, w_2 and w_3 which refer to the weights of the Preact-ResNet50 encoder, the HoVer branch decoder, the NP branch decoder and the NC branch decoder. These 4 sets of weights are optimised jointly using the loss $\mathcal{L}$ defined as:

$$\mathcal{L} = \underbrace{\lambda_a \mathcal{L}_a + \lambda_b \mathcal{L}_b}_{\text{HoVer Branch}} + \underbrace{\lambda_c \mathcal{L}_c + \lambda_d \mathcal{L}_d}_{\text{NP Branch}} + \underbrace{\lambda_e \mathcal{L}_e + \lambda_f \mathcal{L}_f}_{\text{NC Branch}} \tag{3.1}$$

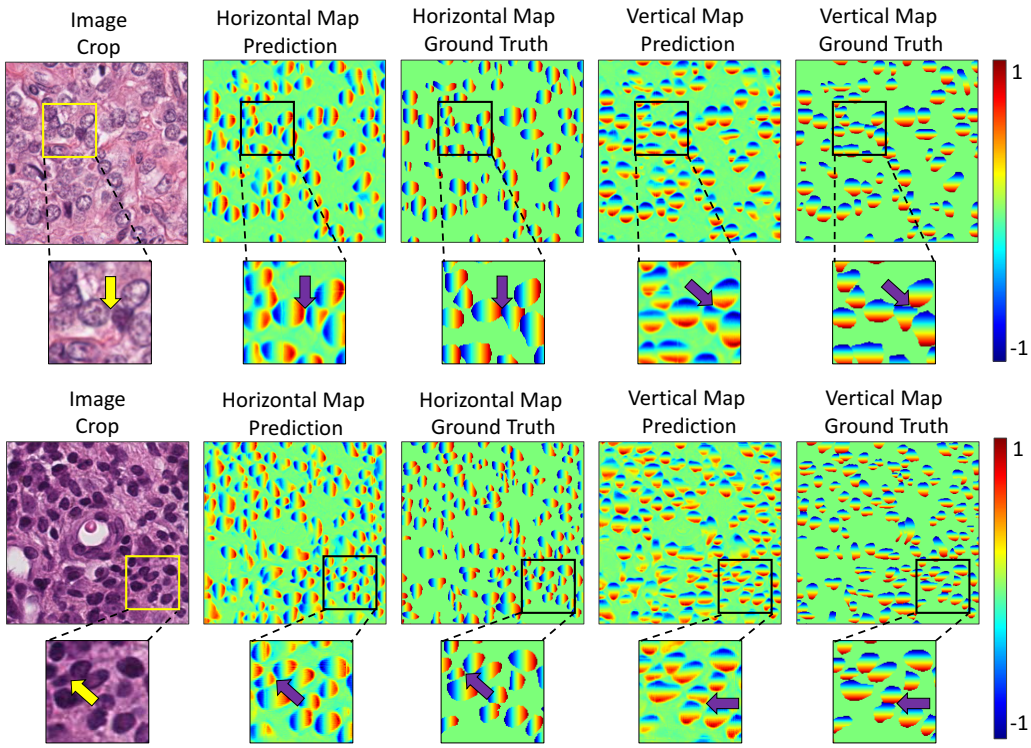


Figure 3.3: Cropped image regions showing horizontal and vertical map predictions, with corresponding ground truth.

where $\mathcal{L}_a$ and $\mathcal{L}_b$ represent the regression loss with respect to the output of the HoVer branch, $\mathcal{L}_c$ and $\mathcal{L}_d$ represent the loss with respect to the output at the NP branch and and finally, $\mathcal{L}_e$ and $\mathcal{L}_f$ represent the loss with respect to the output at the NC branch. We choose to use two different loss functions at the output of each branch for an overall superior performance. $\lambda_a ... \lambda_f$ are scalars that give weight to each associated loss function. Specifically, we set λ_b to 2 and the other scalars to 1, based on empirical selection.

Given the input image $\boldsymbol{x}$, at each pixel i we define $p_i(\boldsymbol{x}, w_0, w_1)$ as the regression output of the HoVer branch, whereas $q_i(\boldsymbol{x}, w_0, w_2)$ and $r_i(\boldsymbol{x}, w_0, w_3)$ denote the pixel-based softmax predictions of the NP and NC branches respectively. We also define $\mathit{\Gamma}_i(\boldsymbol{x})$, $\Psi_i(\boldsymbol{x})$ and $\Phi_i(\boldsymbol{x})$ as their corresponding ground truth (GT). $\Psi_i(\boldsymbol{x})$ is the GT of the nuclear binary map, where background pixels have the value of 0 and nuclear pixels have the value 1. On the other hand, $\Phi_i(\boldsymbol{x})$ is the nuclear type GT where background pixels have the value 0 and any integer value larger than 0 indicates the type of nucleus. Meanwhile, $\mathit{\Gamma}_i(\boldsymbol{x})$ denotes the GT of the horizontal and vertical distances of nuclear pixels to their corresponding centres of mass. For $\mathit{\Gamma}_i(\boldsymbol{x})$, we assign values between -1 and 1 to nuclear pixels in both the horizontal

and vertical directions. We assign the value of the background and the line crossing the centre of mass within each nucleus to be 0. For clarity, we denote the horizontal and vertical components of the GT HoVer map as horizontal map $\Gamma_{i,x}$ and vertical map $\Gamma_{i,y}$ respectively. Visual examples of the horizontal and vertical maps can be seen in Fig. 3.3.

At the output of the HoVer branch, we compute a multiple term regression loss. We denote $\mathcal{L}_a$ as the mean squared error between the predicted horizontal and vertical distances and the GT. We also propose a novel loss function $\mathcal{L}_b$ that calculates the mean squared error between the horizontal and vertical gradients of the horizontal and vertical maps respectively and the corresponding gradients of the GT. We formally define $\mathcal{L}_a$ and $\mathcal{L}_b$ as:

$$\mathcal{L}_a = \frac{1}{N}\sum_{i=1}^{N}(p_i(\boldsymbol{x};\boldsymbol{w}_0,\boldsymbol{w}_1) - \Gamma_i(\boldsymbol{x}))^2 \tag{3.2}$$

$$\begin{aligned}\mathcal{L}_b = &\frac{1}{M}\sum_{i\in\hat{M}}(\nabla_x(p_{i,x}(\boldsymbol{x};\boldsymbol{w}_0,\boldsymbol{w}_1)) - \nabla_x(\Gamma_{i,x}(\boldsymbol{x})))^2 \\ &+\frac{1}{M}\sum_{i\in\hat{M}}(\nabla_y(p_{i,y}(\boldsymbol{x};\boldsymbol{w}_0,\boldsymbol{w}_1)) - \nabla_y(\Gamma_{i,y}(\boldsymbol{x})))^2\end{aligned} \tag{3.3}$$

Within (3.3), ∇_x and ∇_y denote the gradient in the horizontal x and vertical y directions respectively. M denotes total number of nuclear pixels within the image and $\hat{M}$ denotes the set containing all nuclear pixels.

At the output of NP and NC branches, we calculate the cross-entropy loss ($\mathcal{L}_c$ and $\mathcal{L}_e$) and the dice loss ($\mathcal{L}_d$ and $\mathcal{L}_f$). These two losses are then added together to give the overall loss of each branch. Concretely, we define the cross entropy and dice losses as:

$$\text{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} X_{i,k}(\boldsymbol{x})\log Y_{i,k}(\boldsymbol{x}) \tag{3.4}$$

$$\text{Dice} = 1 - \frac{2\times\sum_{i=1}^{N}(Y_i(\boldsymbol{x})\times X_i(\boldsymbol{x})) + \epsilon}{\sum_{i=1}^{N}Y_i(\boldsymbol{x}) + \sum_{i=1}^{N}X_i(\boldsymbol{x}) + \epsilon} \tag{3.5}$$

where X is the ground truth, Y is the prediction, K is the number of classes and ϵ is a smoothness constant which we set to $1.0e^{-3}$. When calculating $\mathcal{L}_c$ and $\mathcal{L}_d$ for NP branch, for a given pixel i, we set X_i and Y_i as $\Psi_i(\boldsymbol{x})$ and $q_i(\boldsymbol{x}, w_0, w_2)$ respectively. For $\mathcal{L}_c$, we set K to be 2 within (3.4) because the task of the branch is to perform binary nuclear segmentation. Similarly, for $\mathcal{L}_e$ and $\mathcal{L}_f$ at the NC branch, for a given pixel i, we substitute X_i for $\Phi_i(\boldsymbol{x})$ and Y_i for $r_i(\boldsymbol{x}, w_0, w_3)$ in (3.4) and (3.5). K is set as 5 within (3.4) when calculating $\mathcal{L}_e$, denoting the 4 types of nuclei that our

model currently predicts and the background. Note, the value of K is chosen to reflect the number of nuclear types represented in the training set.

It must be noted that the NC branch loss $\mathcal{L}_e$ and $\mathcal{L}_f$ are only calculated when the classification labels are available. In other words, as mentioned in Section 3.2.1, the network performs only instance segmentation if there are no classification labels given.

3.2.2 Post Processing

Within each horizontal and vertical map, pixels between separate instances have a significant difference. This can be seen in Fig. 3.3 and is highlighted by the arrows. Therefore, calculating the gradient can inform where the nuclei should be separated because the output will give high values between neighbouring nuclei, where there is a significant difference in the pixel values. We define:

$$\mathcal{S}_m = max(H_x(p_x), H_y(p_y)) \tag{3.6}$$

where p_x and p_y refer to the horizontal and vertical predictions at the output of the HoVer branch and H_x and H_y refer to the horizontal and vertical components of the Sobel operator. Specifically, H_x and H_y compute the horizontal and vertical derivative approximations and are shown by the gradient maps in Fig. 3.1. Therefore, $\mathcal{S}_m$ highlights areas where there is a significant difference in neighbouring pixels within the horizontal and vertical maps. Therefore, areas such as the ones shown by the arrows in Fig. 3.3 will result in high values within $\mathcal{S}_m$. We compute markers $M = \sigma(\tau(q, h) - \tau(S_m, k))$. Here, $\tau(a, b)$ is a threshold function that acts on a and sets values above b to 1 or 0 otherwise. Specifically, h and k were chosen such that they gave the optimal nuclear segmentation results. σ is a rectifier that sets all negative values to 0 and q is the probability map output of the NP branch. We obtain the energy landscape $E = [1 - \tau(S_m, k)] * \tau(q, h)$. Finally, M is used as the marker during marker-controlled watershed to determine how to split $\tau(q, h)$, given the energy landscape E. This sequence of events can be seen in Fig. 3.1.

To perform simultaneous nuclear instance segmentation and classification, it is necessary to convert the per-pixel nuclear type prediction at the output of the NC branch to a prediction per nuclear instance. For each nuclear instance, we use *majority class* of the predictions made by the NC branch, i.e., the nuclear type of all pixels in an instance is assigned to be the class with the highest frequency count for that nuclear instance.

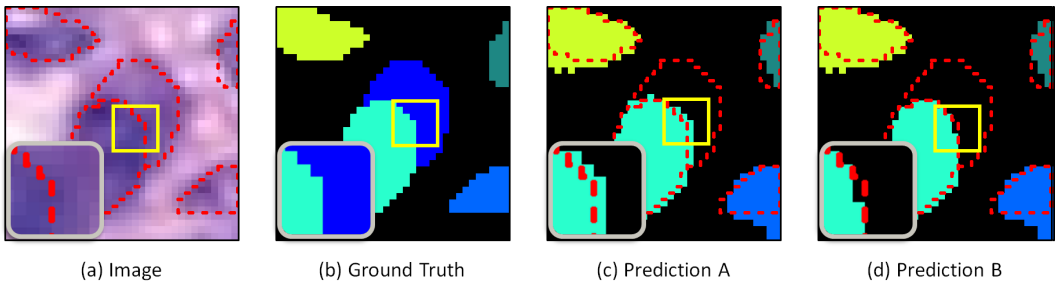


Figure 3.4: Examples highlighting the limitations of DICE2 and AJI with slightly different predictions.

3.3 Evaluation Metrics

3.3.1 Nuclear Instance Segmentation Evaluation

Assessment and comparison of different methods is usually given by an overall score that indicates which method is superior. However, to further investigate the method, it is preferable to break the problem into sub-tasks and measure the performance of the method on each sub-task. This enables an in depth analysis, thus facilitating a comprehensive understanding of the approach, which can help drive forward model development. For nuclear instance segmentation, the problem can be divided into the following three sub-tasks:

- Separate the nuclei from the background
- Detect individual nuclear instances
- Segment each detected instance

In the current literature, two evaluation metrics have been mainly adopted to quantitatively measure the performance of nuclear instance segmentation: 1) Ensemble Dice (DICE2) [151], and 2) Aggregated Jaccard Index (AJI) [88]. Given the ground truth X and prediction Y, DICE2 computes and aggregates DICE per nucleus, where Dice coefficient (DICE) is defined as $2\times(X\cap Y)/(|X|+|Y|)$ and AJI computes the ratio of an aggregated intersection cardinality and an aggregated union cardinality between X and Y.

These two evaluation metrics only provide an overall score for the instance segmentation quality and therefore provides no further insight into the sub-tasks at hand. In addition, these two metrics have a limitation, which we illustrate in Fig. 3.4. From the figure, although prediction A only differs from prediction B by a few pixels, the DICE2 and AJI scores for B are superior. These scores are

Table 3.1: Comparison between Prediction A and Prediction B from Fig.3.4 for various measurements.

	DICE2	**AJI**	**PQ**
Prediction A	0.6477	0.4790	0.6803
Prediction B	0.9007	0.6414	0.6863

shown in Table 3.1. This problem arises due to over-penalisation of the overlapping regions. By overlaying the GT segment contours (red dashed line) upon the two predictions, we observe that, although the cyan-coloured instance within prediction A overlaps mostly with the cyan-coloured GT instance, it also slightly overlaps with the blue-coloured GT instance. As a result, according to the DICE2 algorithm, the predicted cyan instance will be penalised by pixels not only coming from the dominant overlapping cyan-coloured GT instance, but also from the blue-coloured GT instance. The AJI also suffers from the same phenomenon. However, because AJI only uses the prediction and GT instance pair with the highest intersection over union, over-penalisation is less likely compared to DICE2. Over-penalisation is likely to occur when the model completely fails to detect the neighbouring instance, such as in Fig. 3.4. Nonetheless, when evaluating methods across different datasets, specifically on samples containing lots of hard to recognise nuclei such as fibroblasts or nuclei with poor staining, the number of failed detections may increase and therefore may have a negative impact on the AJI measurement. Due to the limitations of DICE2 and AJI, it is clear that there is a need for an improved reliable quantitative measurement.

Panoptic Quality: We propose to use another metric for accurate quantification and interpretability to assess the performance of nuclear instance segmentation. Originally proposed by [83], panoptic quality (PQ) for nuclear instance segmentation is defined as:

$$\mathcal{PQ} = \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{Detection Quality(DQ)}} \times \underbrace{\frac{\sum_{(x,y)\in TP} IoU(x,y)}{|TP|}}_{\text{Segmentation Quality(SQ)}} \tag{3.7}$$

where x denotes a GT segment, y denotes a prediction segment and IoU denotes intersection over union. Each (x,y) pair is mathematically proven to be *unique* [83] over the entire set of prediction and GT segments if their IoU(x,y)>0.5. The unique matching splits all available segments into matched pairs (TP), unmatched GT segments (FN) and unmatched prediction segments (FP). From this, PQ can be intuitively analysed as follows: the *detection quality* (DQ) is the F_1 Score that is

widely used to evaluate instance detection, while *segmentation quality* (SQ) can be interpreted as how close each correctly detected instance is to their matched GT. DQ and SQ, in a way, also provide a direct insight into the second and third sub-tasks, defined above. We believe that PQ should set the standard for measuring the performance of nuclear instance segmentation methods.

Overall, to fully characterise and understand the performance of each method, we use the following three metrics: 1) DICE to measure the separation of all nuclei from the background; 2) Panoptic Quality as a unified score for comparison and 3) AJI for direct comparison with previous publications[3]. Panoptic quality is further broken down into DQ and SQ components for interpretability. Note, SQ is calculated only within true positive segments and should therefore be observed together with DQ. Throughout this study, these metrics are calculated for each image and the average of all images are reported as final values for each dataset.

3.3.2 Nuclear Classification Evaluation

Classification of the type of each nucleus is performed within the nuclear instances extracted from the instance segmentation or detection tasks. Therefore, the overall measurement for nuclear type classification should also encompass these two tasks. For all nuclear instances of a particular type t from both the ground truth and the prediction, the detection task d splits the GT and predicted instances into the following subsets: correctly detected instances (TP_d), misdetected GT instances (FN_d) and overdetected predicted instances (FP_d). Subsequently, the classification task c further breaks TP_d into correctly classified instances of type t (TP_c), correctly classified instances of types other than type t (TN_c), incorrectly classified instances of type t (FP_c) and incorrectly classified instances of types other than type t (FN_c). We then define the F_c score of each type t for combined nuclear type classification and detection as follows:

$$F_c^t = \frac{2(TP_c + TN_c)}{2(TP_c + TN_c) + \alpha_0 FP_c + \alpha_1 FN_c + \alpha_2 FP_d + \alpha_3 FN_d} \tag{3.8}$$

where we use $\alpha_0 = \alpha_1 = 2$ and $\alpha_2 = \alpha_3 = 1$ to give more emphasis to nuclear type classification. Moreover, using the same weighting, if we further extend t to encompass all types of nuclei T ($t \in T$), the classification within TP_d is then divided into a correctly classified set A_c and an incorrectly classified set B_c. We can therefore

[3]Evaluation code available at: `https://github.com/vqdang/hover\_net/src/metrics`

Table 3.2: Summary of the datasets used in our experiments. *Seg* denotes segmentation masks and *Class* denotes classification labels.

	CoNSeP	**Kumar**	**CPM-15**	**CPM-17**	**TNBC**	**CRCHisto**
Number of Nuclei	24,319	21,623	2,905	7,570	4,056	29,756
Labelled Nuclei	24,319	0	0	0	0	22,444
Number of Images	41	30	15	32	50	100
Origin	UHCW	TCGA	TCGA	TCGA	Curie Institute	UHCW
Magnification	40×	40×	40× & 20×	40× & 20×	40×	20×
Size of Images	1000×1000	1000×1000	400×400 to 1000×600	500×500 to 600×600	512×512	500×500
Seg/Class	*Both*	*Seg*	*Seg*	*Seg*	*Seg*	*Class*
Cancer Types	1	8	2	4	1	1

disassemble F_c^t into:

$$\begin{aligned} F_c^T &= \frac{2A_c}{2(A_c+B_c)+FP_d+FN_d} = \frac{2(A_c+B_c)}{2(A_c+B_c)+FP_d+FN_d} \times \frac{A_c}{A_c+B_c} \\ &= F_d \times \text{Classification Accuracy within Correctly Detected Instances} \end{aligned} \tag{3.9}$$

where F_d is simply the standard detection quality like DQ while the other term is the accuracy of nuclear type classification within correctly detected instances. In the case where the GT is not exhaustively annotated for nuclear type classification, like in CRCHisto, an amount equal to the number of unlabelled GT instances in each set is subtracted from B_c and FN_c.

Finally, while IoU is utilised as the criteria in DQ for selecting the TP for detection in instance segmentation, detection methods can not calculate the IoU. Therefore, to facilitate comparison of both instance segmentation and detection methods for the nuclear type classification tasks, for F_c^t, we utilise the notion of distance to determine whether nuclei have been detected. To be precise, we define the region within a predefined radius from the annotated centre of the nucleus as the ground truth and if a prediction lies within this area, then it is considered to be a true positive. Here, we are consistent with [134] and use a radius of 6 pixels at 20× or 12 pixels at 40×.

3.4 Experiments and Results

3.4.1 Datasets

As part of this work, we introduce a new dataset that we term as the colorectal nuclear segmentation and phenotypes (CoNSeP) dataset[4], consisting of 41 H&E stained image tiles, each of size 1,000×1,000 pixels at 40× objective magnification.

[4]This dataset is available at `https://warwick.ac.uk/fac/sci/dcs/research/tia/data/`.

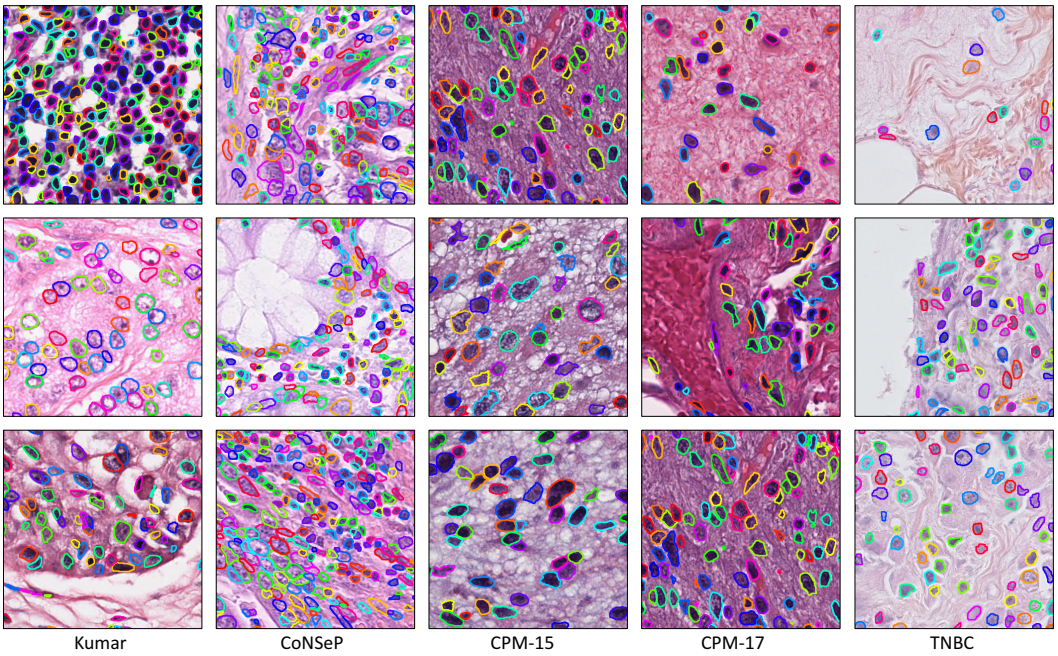


Figure 3.5: Sample cropped regions extracted from each of the five nuclear instance segmentation datasets used in our experiments.

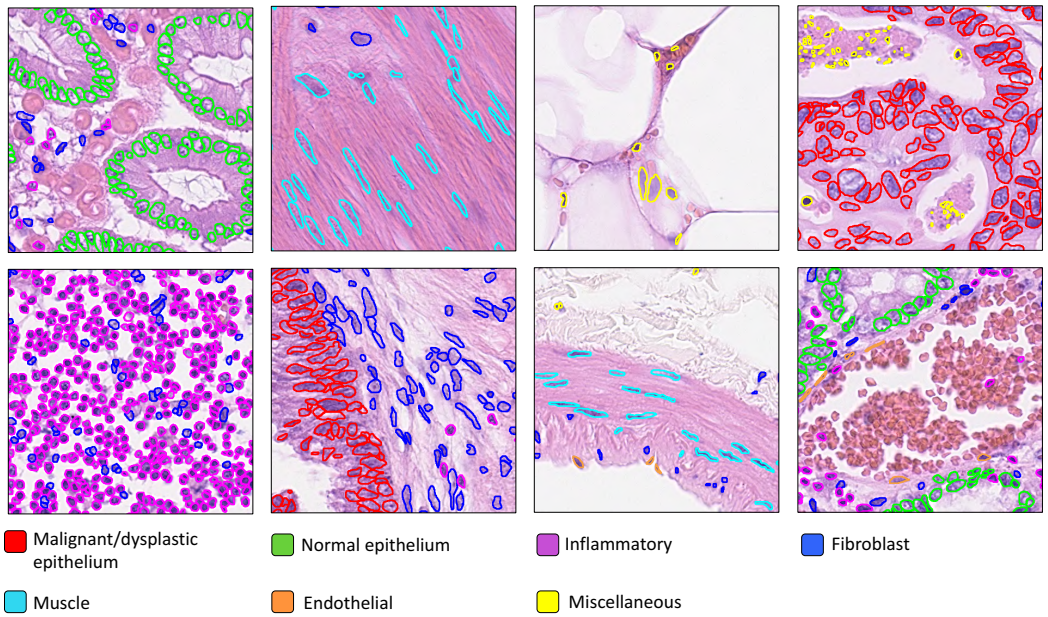


Figure 3.6: Sample cropped regions extracted from the CoNSeP dataset, where the colour of each nuclear boundary denotes the category.

Images were extracted from 16 colorectal adenocarcinoma (CRA) WSIs, each belonging to an individual patient, and scanned with an Omnyx VL120 scanner within

the department of pathology at University Hospitals Coventry and Warwickshire, UK. We chose to focus on a *single* cancer type, so that we are able to display the true variation of tissue within colorectal adenocarcinoma WSIs, as opposed to other datasets that instead focus on using a small number of visual fields from various cancer types. Within this dataset, stroma, glandular, muscular, collagen, fat and tumour regions can be observed. Beside incorporating different tissue components, the 41 images were also chosen such that different nuclei *types* were present, including: normal epithelial; tumour epithelial; inflammatory; necrotic; muscle and fibroblast. Here, by *type* we are referring to the type of cell from which the nucleus originates from. Within the dataset, there are many significantly overlapping nuclei with indistinct boundaries and there exists various artefacts, such as ink. As a result of the diversity of the dataset, it is likely that a model trained on CoNSeP will perform well for unseen CRA cases. For each image tile, every nucleus was annotated by one of two expert pathologists (A.A, Y-W.T). After full annotation, each annotated sample was reviewed by *both* of the pathologists; therefore refining their own and each others' annotations. By the end of the annotation process, each pathologist had fully checked *every* sample and consensus had been reached. Annotating the data in this way ensured that minimal nuclei were missed in the annotation process. However, we can not avoid inevitable pixel-level differences between the annotation and the true nuclear boundary in challenging cases. In addition to delineating the nuclear boundaries, every nucleus was labelled as either: normal epithelial, malignant/dysplastic epithelial, fibroblast, muscle, inflammatory, endothelial or miscellaneous. Within the miscellaneous category, necrotic, mitotic and cells that couldn't be categorised were grouped. For our experiments, we grouped the normal and malignant/dysplastic epithelial nuclei into a single class and we grouped the fibroblast, muscle and endothelial nuclei into a class named spindle-shaped nuclei.

Overall, six independent datasets are utilised for this study. A full summary for each of them is provided in Table 3.2. Five of these datasets are used to evaluate the instance segmentation performance which we refer to as: CoNSeP; Kumar [88]; CPM-15; CPM-17 [151] and TNBC [113]. Example images from each of the five datasets can be seen in Fig. 3.7. Meanwhile, we utilise CoNSeP and a further dataset, named CRCHisto, to quantify the performance of the nuclear classification model. The CRCHisto dataset consists of the same nuclei types that are present in CoNSeP. It is also worth noting that the CRCHisto dataset is not exhaustively annotated for nuclear class labels.

Table 3.3: Comparative experiments on the Kumar [88], CoNSeP and CPM-17 [151] datasets. WS denotes watershed-based post processing.

	Kumar					CoNSeP					CPM-17				
Method	**DICE**	**AJI**	**DQ**	**SQ**	**PQ**	**DICE**	**AJI**	**DQ**	**SQ**	**PQ**	**DICE**	**AJI**	**DQ**	**SQ**	**PQ**
Cell Profiler [24]	0.623	0.366	0.423	0.704	0.300	0.434	0.202	0.249	0.705	0.179	0.570	0.338	0.368	0.702	0.261
QuPath [18]	0.698	0.432	0.511	0.679	0.351	0.588	0.249	0.216	0.641	0.151	0.693	0.398	0.320	0.717	0.230
FCN8 [103]	0.797	0.281	0.434	0.714	0.312	0.756	0.123	0.239	0.682	0.163	0.840	0.397	0.575	0.750	0.435
FCN8 + WS [103]	0.797	0.429	0.590	0.719	0.425	0.758	0.226	0.320	0.676	0.217	0.840	0.397	0.575	0.750	0.435
SegNet [17]	0.811	0.377	0.545	0.742	0.407	0.796	0.194	0.371	0.727	0.270	0.857	0.491	0.679	0.778	0.531
SegNet + WS [17]	0.811	0.508	0.677	0.744	0.506	0.793	0.330	0.464	0.721	0.335	0.856	0.594	0.779	0.784	0.614
U-Net [121]	0.758	0.556	0.691	0.690	0.478	0.724	0.482	0.488	0.671	0.328	0.813	0.643	0.778	0.734	0.578
Mask-RCNN [65]	0.760	0.546	0.704	0.720	0.509	0.740	0.474	0.619	0.740	0.460	0.850	0.684	0.848	0.792	0.674
DCAN [27]	0.792	0.525	0.677	0.725	0.492	0.733	0.289	0.383	0.667	0.256	0.828	0.561	0.732	0.740	0.545
Micro-Net [119]	0.797	0.560	0.692	0.747	0.519	0.794	0.527	0.600	0.745	0.449	0.857	0.668	0.836	0.788	0.661
DIST [113]	0.789	0.559	0.601	0.732	0.443	0.804	0.502	0.544	0.728	0.398	0.826	0.616	0.663	0.754	0.504
CNN3 [88]	0.762	0.508	-	-	-	-	-	-	-	-	-	-	-	-	-
CIA-Net [169]	0.818	**0.620**	0.754	0.762	0.577	-	-	-	-	-	-	-	-	-	-
DRAN [151]	-	-	-	-	-	-	-	-	-	-	0.862	0.683	0.811	0.804	0.657
HoVer-Net	**0.826**	0.618	**0.770**	**0.773**	**0.597**	**0.853**	**0.571**	**0.702**	**0.778**	**0.547**	**0.869**	**0.705**	**0.854**	**0.814**	**0.697**

3.4.2 Implementation and Training Details

We implemented our framework with the open source software library TensorFlow version 1.8.0 [7] on a workstation equipped with two NVIDIA GeForce 1080 Ti GPUs. During training, data augmentation including flip, rotation, Gaussian blur and median blur was applied to all methods. All networks received an input patch with a size ranging from 252×252 to 270×270. This size difference is due to the use of valid convolutions in some architectures, such as HoVer-Net and U-Net. Regarding HoVer-Net, we initialised the model with pre-trained weights on the ImageNet dataset [41], trained only the decoders for the first 50 epochs, and then fine-tuned all layers for another 50 epochs. We train stage one for around 120 minutes and stage two for around 260 minutes. Therefore, the overall training time is around 380 minutes. Stage two takes longer to train because unfreezing the encoder utilises more memory and therefore a smaller batch size needs to be used. Specifically, we used a batch size of 8 and 4 on each GPU for stage one and two respectively. We used Adam optimisation with an initial learning rate of 10^{-4} and then reduced it to a rate of 10^{-5} after 25 epochs. This strategy was repeated for fine-tuning. On the whole, training of the network is stable, where the usage of fully independent decoders helps the network to converge each time. The network was trained with an RGB input, normalised between 0 and 1.

3.4.3 Comparative Analysis of Segmentation Methods

Experimental Setting: We evaluated our approach by employing a full independent comparison across the three largest known exhaustively labelled nuclear

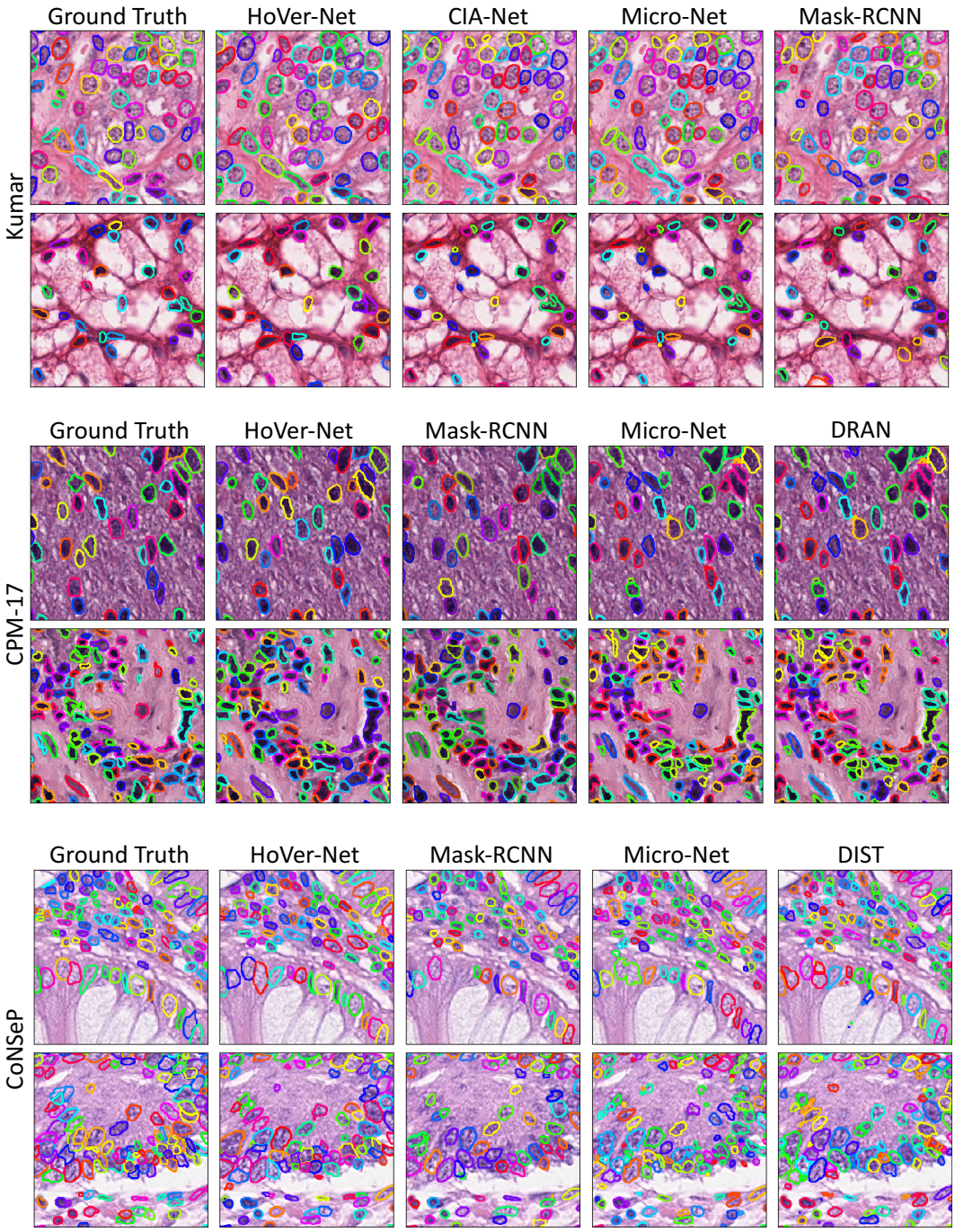


Figure 3.7: Example visual results on the CPM-17, Kumar and CoNSeP datasets. For each dataset, we display the 4 models that achieve the highest PQ score.

segmentation datasets: Kumar; CoNSeP and CPM-17 and utilised the metrics as described in Section 3.3.1. For this experiment, because we do not have the classification labels for all datasets, we perform instance segmentation without classification.

This enables us to fully leverage all data and allows us to rigorously evaluate the segmentation capability of our model. In the same way as [88], we split the Kumar dataset into two different sub-datasets: (i) Kumar-Train, a training set with 16 image tiles (4 breast, 4 liver, 4 kidney and 4 prostate) and (ii) Kumar-Test, a test set with 14 image tiles (2 breast, 2 liver, 2 kidney and 2 prostate, 2 bladder, 2 colon, 2 stomach). Note, we utilise the exact same image split used by other recent approaches [88, 113, 169], but we do not separate the test set into two subsets. We do this to ensure that the test set is large enough, ensuring a reliable evaluation. For CoNSeP, we devise a suitable train and test set that contains 26 and 14 images respectively. The images within the test set were selected to ensure the true diversity of nuclei types within colorectal tissue are represented. For CPM-17, we utilise the same split that had been employed for the challenge, with 32 images in both the training and test datasets.

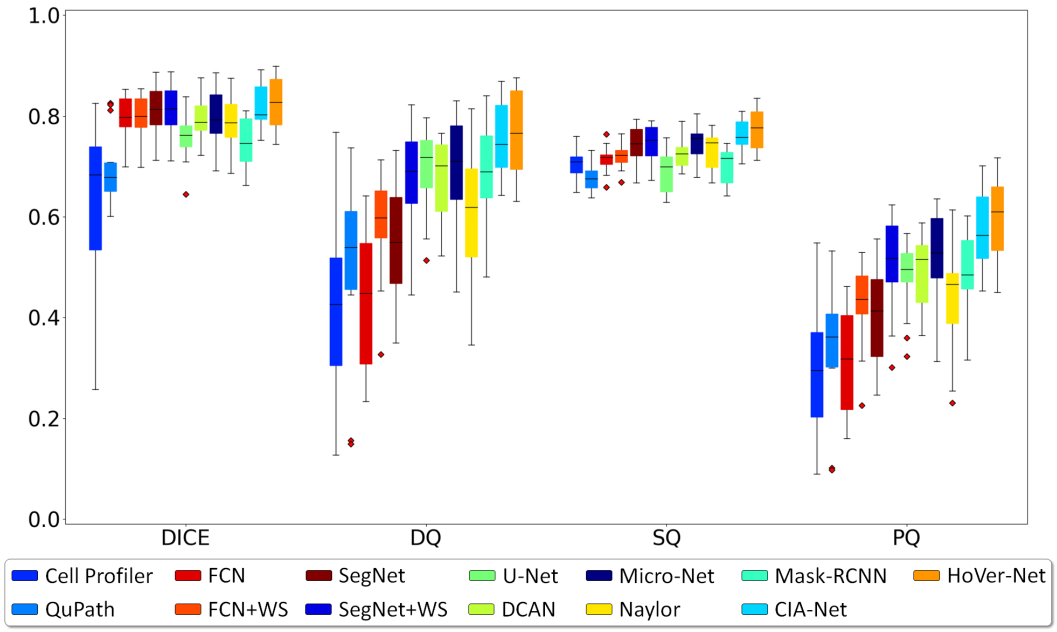
We compared our proposed model to recent segmentation approaches used in computer vision [103, 17, 65], medical imaging [121] and also to methods specifically tuned for the task of nuclear segmentation [27, 119, 113, 169, 151]. We also compared the performance of our model to two open source software applications: Cell Profiler [24] and QuPath [18]. Cell Profiler is a software for cell-based analysis, with several suggested pipelines for computational pathology. The pipeline that we adopted applies a threshold to the greyscale image and then uses a series of post processing operations. QuPath is an open source software for digital pathology and whole-slide image analysis. To achieve nuclear segmentation, we used the default parameters within the application. FCN, SegNet, U-Net, DCAN, Mask-RCNN and DIST have been implemented by the authors of the paper (S.G, Q.D.V). For Mask-RCNN, we slightly modified the original implementation by using smaller anchor boxes. The default configuration is fine-tuned for natural images and therefore, this modification was necessary to perform a successful nuclear segmentation. DIST was implemented with the assistance of the first author of the corresponding approach in order to ensure reliability during evaluation. This also enabled us to utilise DIST for further comparison in our experiments. For Micro-Net, we used the same implementation that was described by [119] and was implemented by the first author of the corresponding paper (S.E.A.R). For CNN3 and CIA-Net, we report the results on the Kumar dataset that are given in their respective original papers. The authors of CIA-Net and DRAN provided their segmentation output, which meant that we were able to obtain all metrics on the datasets that the models were applied to. Therefore, we report results of CIA-Net on the Kumar dataset and results of DRAN on the CPM-17 dataset. Note, for all self-implemented approaches we are

consistent with our pre-processing strategy. However, DRAN, CNN3 and CIA-Net results are directly taken from their respective papers and therefore we can't guarantee the same pre-processing steps. CNN3 and CIA-Net also use stain normalisation, whereas other methods described in this chapter do not.

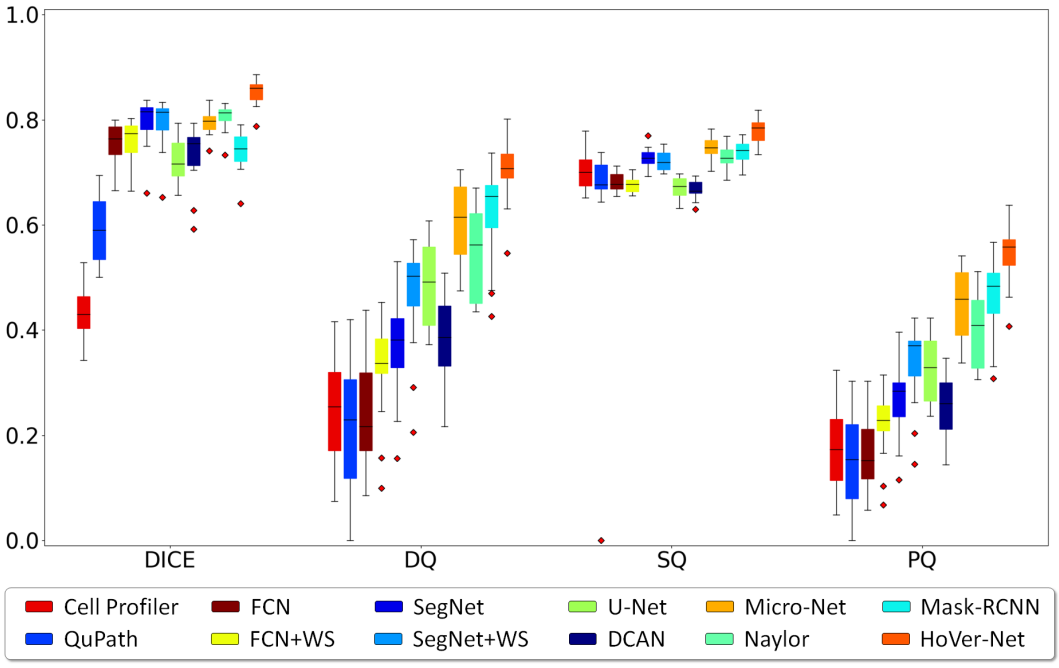
Comparative Results: Table 3.3 and the box plots in Fig. 3.8a and 3.8b show detailed results of this experiment. Within the box plots, we choose not to show AJI, due to its limitations as discussed in Section 3.3.1. A large variation in performance between methods within each dataset is observed. This variation is particularly evident in the Kumar and CoNSeP datasets, where there exists a large number of overlapping nuclei. Both Cell Profiler [24] and QuPath [18] achieve sub-optimal performance for all datasets. In particular, both software applications consistently achieve a low DICE score, suggesting that their inability to distinguish nuclear pixels from the background is a major limiting factor. FCN-based approaches improve the capability of models to detect nuclear pixels, yet often fail due to their inability to separate clustered instances. For example, despite a higher DICE score than Cell Profiler and QuPath, networks built only for semantic segmentation like FCN8 and SegNet suffer from low PQ values. Therefore, methods that incorporate strong instance-aware techniques are favourable. Within CPM-17, there are less overlapping nuclei which explains why methods that are not instance-aware are still able to achieve a satisfactory performance. We observe that the weighted cross entropy loss that is used in both U-Net and Micro-Net can help to separate joined nuclei, but its success also depends on the capacity of the network. This is reflected by the improved performance of Micro-Net over U-Net.

DCAN is able to better distinguish between separate instances than FCN8, which uses a very similar encoder based on the VGG16 network. Therefore, incorporating additional information at the output of the network can improve the segmentation performance. This is also exemplified by the fairly strong performances of CNN3, DIST, DRAN and CIA-Net. In a different way, Mask-RCNN is able to successfully separate clustered nuclei by utilising a region proposal based approach. However, Mask-RCNN is less effective than other methods at detecting nuclear pixels, which is reflected by a lower DICE score.

Due to the reasoning given in Section 3.3, we place a larger emphasis on PQ to determine the success of different models. In particular, we consistently obtain an improved performance over DIST, which justifies the use of our proposed horizontal and vertical maps as a regression target. We also report a better performance than the winners of the Computational Precision Medicine and MoNuSeg challenges [151, 169], that utlised the CPM-17 and Kumar datasets respectively. Therefore,



(a) Kumar



(b) CoNSeP

Figure 3.8: Box plots highlighting the performance of competing methods on the Kumar and CoNSeP datasets.

HoVer-Net achieves state-of-the art performance for nuclear instance segmentation compared to all competing methods on multiple datasets that consist of a variety of

different tissue types. Our approach also outperforms methods that were fine-tuned for the task of nuclear segmentation.

3.4.4 Generalisation Study

Experimental Setting: The goal of any automated method is to perform well on unseen data, with high accuracy. Therefore, we conducted a large scale study to assess how all methods generalise to new H&E stained images. To analyse the generalisation capability, we assessed the ability to segment nuclei from: i) new organs (variation in nuclei shapes) and ii) different centres (variation in staining).

The five instance segmentation datasets used within our experiments can be grouped into three groups according to their origin: TCGA (Kumar, CPM-15, CPM-17), TNBC and CoNSeP. We used Kumar as the training and validation set, due to its size and diversity, whilst the combined CPM (CPM-15 and CPM-17), TNBC and CoNSeP datsets were used as three independent test sets. We split the test sets in this way in accordance with their origin. Note, for this experiment we use both the training and test sets of CPM-17 and CoNSeP to form the independent test sets. Kumar was split into three subsets, as explained in Section 5.4.1, and Kumar-Train was used to train all models, i.e. trained with samples originating from the following organs: breast; prostate; kidney and liver. Despite all samples being extracted from TCGA, CPM samples come from the brain, head & neck and lungs regions. Therefore, testing with CPM reflects the ability for the model to generalise to new organs, as mentioned above by the first generalisation criterion. TNBC contains samples from an already seen organ (breast), but the data is extracted from an independent source with different specimen preservation and staining practice. Therefore, this reflects the second generalisation criterion. CoNSeP contains samples taken from colorectal tissue, which is not represented in Kumar-Train, and is also extracted from a source independent to TCGA. Therefore, this reflects *both* the first and second generalisation criteria. Also, as mentioned in Section 5.4.1, CoNSeP contains challenging samples, where there exists various artefacts and there is variation in the quality of slide preparation. Therefore, the performance on this dataset also reflects the ability of a model to generalise to difficult samples.

Comparative Results: The results are reported in Table 3.4, where we only display the results of methods that employ an instance-based technique. We observe that our proposed model is able to successfully generalise to unseen data in all three cases. However, some methods prove to perform poorly with unseen data, where in particular, U-Net and DIST perform worse than other competing methods on all three datasets. Both SegNet with watershed and Mask-RCNN achieve a competitive

Table 3.4: Generalisation capability of different models for nuclear segmentation. All models are initially trained on Kumar and then the Combined CPM [151], TNBC [113] and CoNSeP datasets are processed.

	Combined CPM					TNBC					All CoNSeP				
Method	**DICE**	**AJI**	**DQ**	**SQ**	**PQ**	**DICE**	**AJI**	**DQ**	**SQ**	**PQ**	**DICE**	**AJI**	**DQ**	**SQ**	**PQ**
FCN8 + WS [103]	0.762	0.531	0.669	0.722	0.487	0.726	0.506	0.662	0.723	0.480	0.609	0.247	0.345	0.688	0.240
SegNet + WS [17]	0.791	0.583	0.738	0.755	0.561	0.758	0.559	0.734	0.750	0.554	0.681	0.315	0.449	0.733	0.332
U-Net [121]	0.720	0.541	0.652	0.672	0.446	0.681	0.514	0.635	0.676	0.442	0.585	0.363	0.442	0.670	0.297
Mask-RCNN [65]	0.764	0.575	0.760	0.719	0.549	0.705	0.529	0.726	0.742	0.543	0.606	0.348	0.492	0.720	0.357
DCAN [27]	0.770	0.582	0.716	0.730	0.528	0.725	0.537	0.683	0.720	0.495	0.609	0.306	0.403	0.685	0.278
Micro-Net [119]	0.792	0.615	0.716	0.751	0.542	0.701	0.531	0.656	0.753	0.497	0.644	0.394	0.489	0.722	0.356
DIST [113]	0.775	0.563	0.593	0.720	0.432	0.719	0.523	0.549	0.714	0.404	0.621	0.369	0.379	0.701	0.268
HoVer-Net	**0.801**	**0.626**	**0.774**	**0.778**	**0.606**	**0.749**	**0.590**	**0.743**	**0.759**	**0.578**	**0.664**	**0.404**	**0.529**	**0.764**	**0.408**

performance across all three generalisation tests. However, similar to the results reported in Table 3.3, Mask-RCNN is not able to distinguish nuclear pixels from the background as well as other competing methods, which has an adverse effect on the overall segmentation performance shown by PQ. On the other hand, SegNet proves to successfully detect nuclear pixels, reporting a greater DICE score than HoVer-Net on both the TNBC and CoNSeP datasets. However, the overall segmentation result for HoVer-Net is superior because it is better able to separate nuclear instances by incorporating the horizontal and vertical maps at the output of the network.

3.4.5 Comparative Analysis of Classification Methods

Experimental Setting: We converted the top four performing nuclear instance segmentation algorithms, based on their panoptic quality on the CoNSeP dataset, such that they were able to perform simultaneous instance segmentation and classification. As mentioned in Section 5.4.1, the nuclear categories that we use in our experiments are: miscellaneous, inflammatory, epithelial and spindle-shaped. Specifically, we compared HoVer-Net with Micro-Net, Mask-RCNN and DIST. For Micro-Net, we used an output depth of 5 rather than 2, where each channel gave the probability of a pixel being either background, miscellaneous, inflammatory, epithelial or spindle-shaped. For Mask-RCNN, there is a devoted classification branch that predicts the class of each instance and therefore is well suited to a multi-class setting. DIST performs regression at the output of the network and therefore converting the model such that it is able to classify nuclei into multiple categories is non-trivial. Instead, we add an extra 1×1 convolution at the output of the network that performs nuclear classification. As well as comparing to the aforementioned methods, we compared our approach to a spatially constrained CNN (SC-CNN), that achieves detection and classification. Note, because SC-CNN does not produce a segmentation mask, we do not report the PQ for this method.

Comparative Results: We trained our models on the training set of the CoNSeP dataset and then we evaluated the model on both the test set of CoNSeP and also the entire CRCHisto dataset. Table 3.5 displays the results of the multi-class models on the CoNSeP and the CRCHisto datasets respectively, where the given metrics are described in Section 3.3.2. For CoNSeP, along with the classification metrics, we provide PQ as an indication of the quality of instance segmentation. However, in CRCHisto, only the nuclear centroids are given and therefore, we exclude PQ from the CRCHisto evaluation because it can't be calculated without the instance segmentation masks. We observe that HoVer-Net achieves a good quality simultaneous instance segmentation and classification, compared to competing methods. It must be noted, that we should expect a lower F_1 score for the miscellaneous class because there are significantly less nuclei represented. Also, there is a high diversity of nuclei types that have been grouped within this class, belonging to: mitotic; necrotic and cells that are uncategorisable. Despite this, HoVer-Net is able to achieve a satisfactory performance on this class, where other methods fail. Furthermore, compared to other methods, our approach achieves the best F_1 score for epithelial, inflammatory and spindle classes. Therefore, due to HoVer-Net obtaining a strong performance for both nuclear segmentation and classification, we suggest that our model may be used for sophisticated subsequent cell-level downstream analysis in computational pathology.

Table 3.5: Comparative results for nuclear classification on the CoNSeP and CRCHisto datasets. F_d denotes the F_1 score for nuclear detection, whereas F_c^e, F_c^i, F_c^s and F_c^m denote the F_1 classification score for the epithelial, inflammatory, spindle-shaped and miscellaneous classes.

	CoNSeP						**CRCHisto**				
Method	**PQ**	$\mathbf{F}_d$	$\mathbf{F}_c^e$	$\mathbf{F}_c^i$	$\mathbf{F}_c^s$	$\mathbf{F}_c^m$	$\mathbf{F}_d$	$\mathbf{F}_c^e$	$\mathbf{F}_c^i$	$\mathbf{F}_c^s$	$\mathbf{F}_c^m$
SC-CNN [134]	-	0.608	0.306	0.193	0.175	0.000	0.664	0.246	0.111	0.126	0.000
DIST [113]	0.372	0.712	0.617	0.534	0.505	0.000	0.616	0.464	0.514	0.275	0.000
Micro-Net [119]	0.430	0.743	0.615	0.592	0.532	0.117	0.638	0.422	0.518	0.249	0.059
Mask-RCNN [65]	0.450	0.692	0.595	0.590	0.520	0.098	0.639	**0.503**	0.537	0.294	0.077
HoVer-Net	**0.516**	**0.748**	**0.635**	**0.631**	**0.566**	**0.426**	**0.688**	0.486	**0.573**	**0.302**	**0.178**

3.4.6 Ablation Study

To gain a full understanding of the contribution of our method, we investigated several of its components. Specifically, we performed the following ablation experiments: (i) contribution of the proposed loss strategy; (ii) Sobel-based post processing technique compared to other strategies and (iii) contribution of the dedicated

classification branch. Here, we utilised the Kumar and CoNSeP datasets for (i) and (ii) due to the large number of nuclei present, whereas for (iii) we use CoNSeP and CRCHisto because we do not have the classification labels for Kumar.

Loss Terms: We conducted an experiment to understand the contribution of our proposed loss strategy. First, we used mean squared error (MSE) of the horizontal and vertical distances L_a as the loss function of the HoVer branch and binary cross entropy (BCE) loss L_c as the loss function for the NP branch. We refer to this combination as the *standard* strategy because MSE and BCE are the two most commonly used loss functions for regression and binary classification tasks respectively. Next, we introduced the MSE of the horizontal and vertical gradients L_b to the HoVer branch and the dice loss L_d to the NP branch. The intuition behind our novel L_b is that it enforces the correct structure of the horizontal and vertical map predictions and therefore helps to correctly separate neighbouring instances. The dice loss was introduced because it can help the network to better distinguish between background and nuclear pixels and is particularly useful when there is a class-imbalance. We present the results in Table 3.6, where we observe an increase in all performance measures for our proposed multi-term loss strategy. Therefore, the additional loss terms boost the network's ability to differentiate between nuclear and background pixels (DICE) and separate individual nuclei (DQ and PQ). In particular, there is a significant boost in the SQ for both Kumar and CoNSeP, which suggests that our proposed loss function L_b is necessary to precisely determine where nuclei should be split.

Post Processing: Usually, markers obtained from applying a threshold to an energy landscape (such as the distance map) is enough to provide a competitive input for watershed, as seen by DIST in Table 3.3. Although HoVer-Net is not directly built upon an energy landscape, we devised a Sobel-based method to derive both the energy landscape and the markers. To compare with other methods, we implemented two further techniques for obtaining the energy landscape and the markers. We then exhaustively compared all energy landscape and marker combinations to assess which post processing strategy is the best. We start by linking HoVer to the distance map by calculating the square sum $\chi^2 + \varphi^2$, which can be seen as the distance from a pixel to its nearest nuclear centroid. In other words, this is a pseudo distance map. Additionally, χ and φ values can be interpreted as Cartesian coordinates with each nuclear centroid as the origin. By thresholding the values between a certain range, we can obtain the markers. The results of all combinations are shown in Table 3.7. Note, our gradient-based post processing technique is specifically designed for the HoVer branch output.

Table 3.6: Ablation study highlighting the contribution of the proposed loss strategy.

	Kumar					CoNSeP				
Strategy	**DICE**	**AJI**	**DQ**	**SQ**	**PQ**	**DICE**	**AJI**	**DQ**	**SQ**	**PQ**
Standard Loss	0.823	0.750	0.771	0.581	0.608	0.846	0.685	0.774	0.532	0.557
Proposed Loss	**0.826**	**0.770**	**0.773**	**0.597**	**0.618**	**0.853**	**0.702**	**0.778**	**0.547**	**0.571**

Table 3.7: Ablation study for post processing techniques: Sobel-based versus thresholding to get markers and Sobel-based versus naive conversion to get energy landscape.

		Kumar					CoNSeP				
Energy	**Markers**	**DICE**	**AJI**	**DQ**	**SQ**	**PQ**	**DICE**	**AJI**	**DQ**	**SQ**	**PQ**
$\chi^2 + \varphi^2$	Threshold	0.825	0.597	0.705	0.764	0.541	0.850	0.543	0.602	0.761	0.459
$\chi^2 + \varphi^2$	Sobel	0.826	0.613	0.766	0.768	0.591	0.853	0.561	0.694	0.770	0.535
Sobel	Threshold	0.825	0.614	0.715	0.772	0.554	0.850	0.566	0.617	0.775	0.479
Sobel	Sobel	**0.826**	**0.618**	**0.770**	**0.773**	**0.597**	**0.853**	**0.571**	**0.702**	**0.778**	**0.547**

Classification Branch: In order to assess the importance of a devoted branch for concurrent nuclear segmentation and classification, we compared the proposed three branch setup of HoVer-Net to a two branch setup. Here, the two branch setup extends the NP branch to a multi-class setting, by predicting each nuclear type at the output. Then, to obtain the binary mask, the positive channels are combined together after nuclear type prediction. Utilising three branches decouples the tasks of nuclear classification and nuclear detection, where a separate branch is devoted to each task. For this ablation study, we train on the CoNSeP training set and then process both the CoNSeP test set and the entire CRCHisto dataset.

We report results in Table 3.8, where we observe that utilising a separate branch devoted to the task of nuclear classification leads to an improved overall performance of simultaneous nuclear instance segmentation and classification in both the CoNSeP and CRCHisto datasets. We can see that if the classification takes place at the output of NP branch, then the network's ability to determine the nuclear type is compromised. This is because the task of nuclear classification is challenging and therefore the network benefits from the introduction of a branch dedicated to the task of classification.

3.5 Discussion and Conclusions

Analysis of nuclei in large-scale histopathology images is an important step towards automated downstream analysis for diagnosis and prognosis of cancer. Nuclear fea-

Table 3.8: Ablation study showing the contribution of the HoVer-Net classification branch on the CoNSeP dataset. F_d denotes the F_1 score for nuclear detection, whereas F_c^e, F_c^i, F_c^s and F_c^m denote the F_1 classification score for the epithelial, inflammatory, spindle-shaped and miscellaneous classes.

	CoNSeP						**CRCHisto**				
Branches	**PQ**	$\mathbf{F}_d$	$\mathbf{F}_c^e$	$\mathbf{F}_c^i$	$\mathbf{F}_c^s$	$\mathbf{F}_c^m$	$\mathbf{F}_d$	$\mathbf{F}_c^e$	$\mathbf{F}_c^i$	$\mathbf{F}_c^s$	$\mathbf{F}_c^m$
NP & HoVer	0.499	0.736	**0.636**	0.545	0.528	0.333	0.666	0.458	0.523	0.271	0.132
NP & HoVer & NC	**0.516**	**0.748**	0.635	**0.631**	**0.566**	**0.426**	**0.688**	**0.486**	**0.573**	**0.302**	**0.178**

tures have been often used to assess the degree of malignancy [63]. However, visual analysis of nuclei is a very time consuming task because there are often tens of thousands of nuclei within a given whole-slide image (WSI). Performing simultaneous nuclear instance segmentation and classification enables subsequent exploration of the role that nuclear features play in predicting clinical outcome. For example, [104] utilised nuclear features from histology TMA cores to predict survival in early-stage estrogen receptor-positive breast cancer. Restricting the analysis to some specific nuclear types only may be advantageous for accurate analysis in computational pathology.

In this chapter, we have proposed HoVer-Net for simultaneous segmentation and classification of nuclei within multi-tissue histology images that not only detects nuclei with high accuracy, but also effectively separates clustered nuclei. Our approach has three upsampling branches: 1) the nuclear pixel branch that separates nuclear pixels from the background; 2) the HoVer branch that regresses the horizontal and vertical distances of nuclear pixels to their centres of mass and 3) the nuclear classification branch that determines the type of each nucleus. We have shown that the proposed approach achieves the state-of-the-art instance segmentation performance compared to a large number of recently published deep learning models across multiple datasets, including tissues that have been prepared and stained under different conditions. This makes the proposed approach likely to translate well to a practical setting due its strong generalisation capacity, which can therefore be effectively used as a prerequisite step before nuclear-based feature extraction. We have shown that utilising the horizontal and vertical distances of nuclear pixels to their centres of mass provides powerful instance-rich information, leading to state-of-the-art performance in histological nuclear segmentation. When the classification labels are available, we show that our model is able to successfully segment and classify nuclei with high accuracy.

Despite us extensively validating the superior performance of our model,

there are various shortcomings that may be addressed in future work. For example, the concept of horizontal and vertical maps are better suited to convex objects, such as most nuclei, and therefore may not necessarily translate well to other tasks such as gland segmentation. Future work may involve the development of our horizontal and vertical targets so that they are better suited to general object segmentation in computational pathology. Another disadvantage of our approach is that it assumes that clustered nuclei are generally positioned above/below and side by side, due to the configuration of our HoVer maps and therefore may not perform well when nuclei are positioned at other angles to each other. A natural extension would be to develop the idea of horizontal and vertical maps and include additional directions to those utilised in our approach.

Region proposal (RP) methods, such as Mask-RCNN, show great potential in dealing with overlapping instances because there is no notion of *separating* instances; instead nuclei are segmented independently. However, a major limitation of the RP methods is the difficulty in merging instance predictions between neighbouring tiles during processing. For example, if a sub-segment of a nucleus at the boundary is assigned a label, one must ensure that the remainder of the nucleus in the neighbouring tile is also assigned the same label. To overcome this difficulty, for Mask-RCNN, we utilised an overlapping tile mechanism such that we only considered non-boundary nuclei.

Regarding the processing time, the average time to process a 1,000×1,000 image tile over 10 runs using Mask-RCNN for segmentation and classification was 106.98 seconds. Meanwhile, HoVer-Net only took an average of 11.04 seconds to complete the same operation; approximately 9.7× faster. On the other hand, the average processing time for DIST and Micro-Net was 0.600 and 0.832 seconds respectively. Mask-RCNN inherently stores a single instance per channel, which leads to very large arrays in memory when there are many nuclei in a single image patch, which also contributes to the much longer processing time as seen above. Overall, FCN methods seem to better translate to WSI processing compared to Mask-RCNN or RPN methods in general. It must be stressed that the timing is not exact and is dependent on hardware specifications and software implementation. With optimised code and sophisticated hardware, we expect these timings to be considerably different. Additionally, the inference time is also dependent on the size of the output. In particular, with a smaller output size, a smaller stride is also required during processing. For instance, if we used padded convolution in the upsampling branches of HoVer-Net, then we observe 5.6× speed up and the average processing time is 1.97 seconds per 1000×1000 image tile. For fair comparison, all models were pro-

cessed on a single GPU with 12GB RAM and we fixed the batch size to a size of one. Future work will explore the trade-off between the efficiency of HoVer-Net and its potential to accurately perform instance segmentation and classification.

A major bottleneck for the development of successful nuclear segmentation algorithms is the limitation of data; particularly with additional associated class labels. In this work, we introduce the colorectal adenocarcinoma nuclear segmentation and phenotypes (CoNSeP) dataset, containing over 24K labelled nuclei from challenging samples to reflect the true difficulty of segmenting nuclei in whole-slide images. Due to the abundance of nuclei with an associated nuclear category, CoNSeP aims to help accelerate the development of further simultaneous nuclear instance segmentation and classification models to further increase the sophistication of cell-level analysis within computational pathology.

We analysed the common measurements used to assess the true performance of nuclear segmentation models and discussed their limitations. Due to the fact that these measurements did not always reflect the instance segmentation performance, we proposed a set of reliable and informative statistical measures. We encourage researchers to utilise the proposed measures to not only maximise the interpretability of their results, but also to perform a fair comparison with other methods.

Finally, methods have surfaced recently that explore the relationship of various nuclear types within histology images [74, 136], yet these methods are limited to spatial analysis because the segmentation masks are not available. Utilising our model for nuclear segmentation and classification enables the exploration of the spatial relationship between various nuclear types combined with nuclear morphological features and therefore may provide additional diagnostic and prognostic value. Currently, our model is trained on a single tissue type, yet due to the strong performance of our instance segmentation model across multiple tissues, we are confident that our model will perform well if we were to incorporate additional tissue types. We observe a low F_1 classification score for the miscellaneous category in the classification model because there are significantly less samples within this category and there exists high intra-class variability. Future work will involve obtaining more samples within this category, including necrotic and mitotic nuclei, to improve the class balance of the data.

Chapter 4

MILD-Net for Gland Instance Segmentation

Colorectal cancer is the third most commonly occurring cancer in men and the second most commonly occurring cancer in women, where approximately 95% of all colorectal cancers are adenocarcinomas [46]. Colorectal adenocarcinoma develops in the lining of the colon or rectum, which makes up the large intestine and is characterised by glandular formation. Histological examination of the glands, most frequently with the Hematoxylin & Eosin (H&E) stain, is routine practice for assessing the differentiation of the cancer within colorectal adenocarcinoma. Pathologists use the degree of glandular formation as an important factor in deciding the grade or degree of differentiation of the tumour. Within well differentiated cases, above 95% of the tumour is gland forming [46], whereas in poorly differentiated cases, typical glandular appearance is lost. Within the top row of Figure 4.1, (a) shows a healthy case, (b) shows a moderately differentiated tumour and (c) shows a poorly differentiated tumour. We observe the loss of glandular formation as the grade of cancer increases.

There is a growing trend towards a digitised pathology workflow, where digital images are acquired from glass histology slides using a scanning device. The advent of digital pathology has led to a rise in computational pathology, where algorithms are implemented to assist pathologists in diagnostic decision making. In routine pathological practice, accurate segmentation of structures such as glands and nuclei are of crucial importance because their morphological properties can assist a pathologist in assessing the degree of malignancy [34, 64, 154]. With the advent of computational pathology, digitised histology slides are being leveraged such that pathological segmentation tasks can be completed in an objective manner. In par-

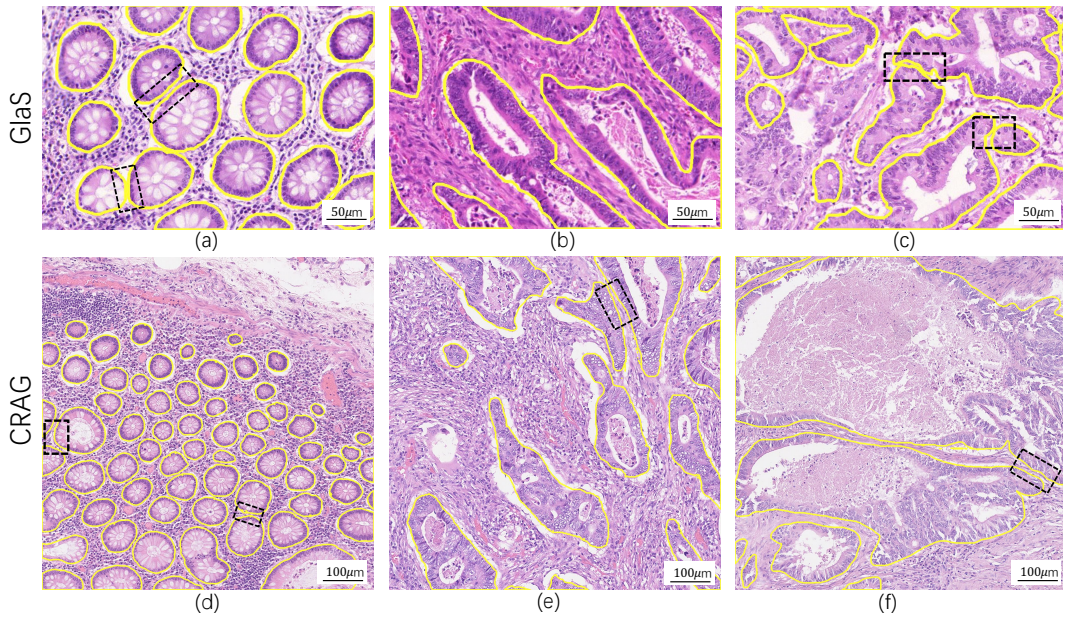


Figure 4.1: (a-c) Example images from the GlaS dataset [135]. (d-f) Example images from the CRAG dataset.

ticular, automated gland segmentation within H&E images can enable pathologists to extract vital morphological features from large scale histopathology images, that would otherwise be impractical.

However, automated gland segmentation remains a challenging task due to several important factors. First, a high resolution level is needed for precise delineation of glandular boundaries, that is important when extracting morphological measurements. Next, glands vary in their size and shape, especially as the grade of cancer increases. Furthermore, the output of solely the gland object gives limited information when making a diagnosis. Extra information, such as the uncertainty of a prediction and the simultaneous segmentation of additional histological components, may give additional diagnostic power. For example, the pathologist may choose to ignore areas with high uncertainty, such as areas with dense nuclei and areas containing artefacts. An additional histological component of particular interest is the lumen, which is ultimately the defining structure of a gland. This structure can help empower diagnostic decision making, because its presence and morphology can be indicative of the grade of cancer.

In this chapter we propose a minimal information loss dilated network that aims to solve the key challenges posed by automated gland segmentation. During feature extraction, we introduce minimal information loss (MIL) units, where we incorporate the original downsampled image into the residual unit after max-pooling.

This, alongside dilated convolution, helps retain maximal information that is essential for segmentation, particularly at the glandular boundaries. We use *atrous* spatial pyramid pooling for multi-scale aggregation that is essential when segmenting glands of varying shapes and sizes. After feature extraction, our network upsamples the feature maps to localise the regions of interest. For uncertainty quantification, we apply random transformations to the input images as a method of generating the predictive distribution. This leads to a superior segmentation result and allows us to observe areas of uncertainty that may be clinically informative. Furthermore, we use this measure of uncertainty to devise a scheme for ranking images to prioritise for pathologist annotation. As an extension, we demonstrate how our method can be modified to simultaneously segment the gland lumen. The additional segmentation of the gland lumen can empower current automated methods to achieve a more accurate diagnosis.

Experimental results show that the proposed framework achieves state-of-the-art performance on the 2015 MICCAI GlaS Challenge dataset and on a second independent colorectal adenocarcinoma dataset.

4.1 Related Work

Computerised techniques play a significant role in automated digitalised histology image analysis, with applications to various tasks including but limited to nuclei detection and segmentation [59, 26, 134], mitosis detection [31, 25, 150, 10], tumour segmentation [117], image retrieval [126, 132], cancer type classification [60, 85, 19, 98, 116], etc. Most of the previous literature focused on the hand-crated features for histopathological image analysis [63]. Recently, deep learning achieved great success on image recognition tasks with powerful feature representation [101, 131, 94]. For example, U-Net achieved excellent performance on the gland segmentation task [121]. To further improve the gland instance segmentation performance, Chen et al. presented a deep contour-aware network by formulating an explicit contour loss function in the training process and achieved the best performance during the 2015 MICCAI Gland Segmentation (GlaS) on-site challenge [27, 26, 135]. In addition, a framework was proposed by Xu *et al.* [160] by fusing complex multichannel regional and boundary patterns with side supervision for gland instance segmentation. This work was extended in [161] to incorporate additional bounding box information for an enhanced performance. A Multi-Input-Multi-Output network (MIMO-Net) was presented for gland segmentation in [118] and achieved the state-of-the-art performance. Furthermore, several methods have investigated the segmentation of glands

from histology images using limited expert annotation effort. For example, a deep active learning framework was presented in [163] for gland segmentation using suggestive annotation. Unannotated images were utilised in [168] with the design of deep adversarial networks and consistently good segmentation performance was attained.

4.2 Methods

4.2.1 Minimal Information Loss Dilated Network

Gland instance segmentation is a complex task that requires a significantly deep network for meaningful feature extraction. Therefore, we use residual units to allow efficient gradient propagation through our deep network architecture. Traditional convolutional neural networks use a combination of max-pooling and convolution in a hierarchical fashion to increase the size of the receptive field [94]. The inclusion of max-pooling results in the loss of information with relatively low activations [125], that is important for pixel-level prediction in segmentation. A significant amount of downsampling via max-pooling leads to a sub-optimal segmentation, particularly at thin object boundaries and for small objects. To counter this loss of information, in addition to using traditional residual units, we include two additional types of residual units during feature extraction: MIL units and dilated residual units. The MIL unit incorporates the original image into each residual unit directly after the max-pooling layer. First, the original image is downsampled to the same size as the output of the pooling operation by bicubic interpolation. Then, a 3×3 convolution is applied before concatenating to the output of the pooling layer. Next, a 3×3 convolution is applied to the concatenated block and this output is subsequently used in the residual summation operation, as opposed to the input tensor in traditional methods. Three MIL units are added during feature extraction immediately after max-pooling. These MIL units can be seen in more detail within part (a) of Figure 4.2. A traditional residual unit, which is defined as:

$$\boldsymbol{y} = \mathcal{F}(\boldsymbol{x}, \boldsymbol{W}) + \boldsymbol{x} \tag{4.1}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ denote the input and output vectors respectively and $\boldsymbol{W}$ denotes the weights within the residual unit. Specifically $\mathcal{F}$ represents the function $\boldsymbol{W}_2(\sigma(\boldsymbol{W}_1\boldsymbol{x}))$, where σ denotes ReLU, $\boldsymbol{W}_1$ denotes the weights of the first convolution and $\boldsymbol{W}_2$ denotes the weights of the second convolution. The addition of the the input vector $\boldsymbol{x}$ to $\mathcal{F}$ is shown by the summation operator $\oplus$ in the residual unit of part (d) in Fig-

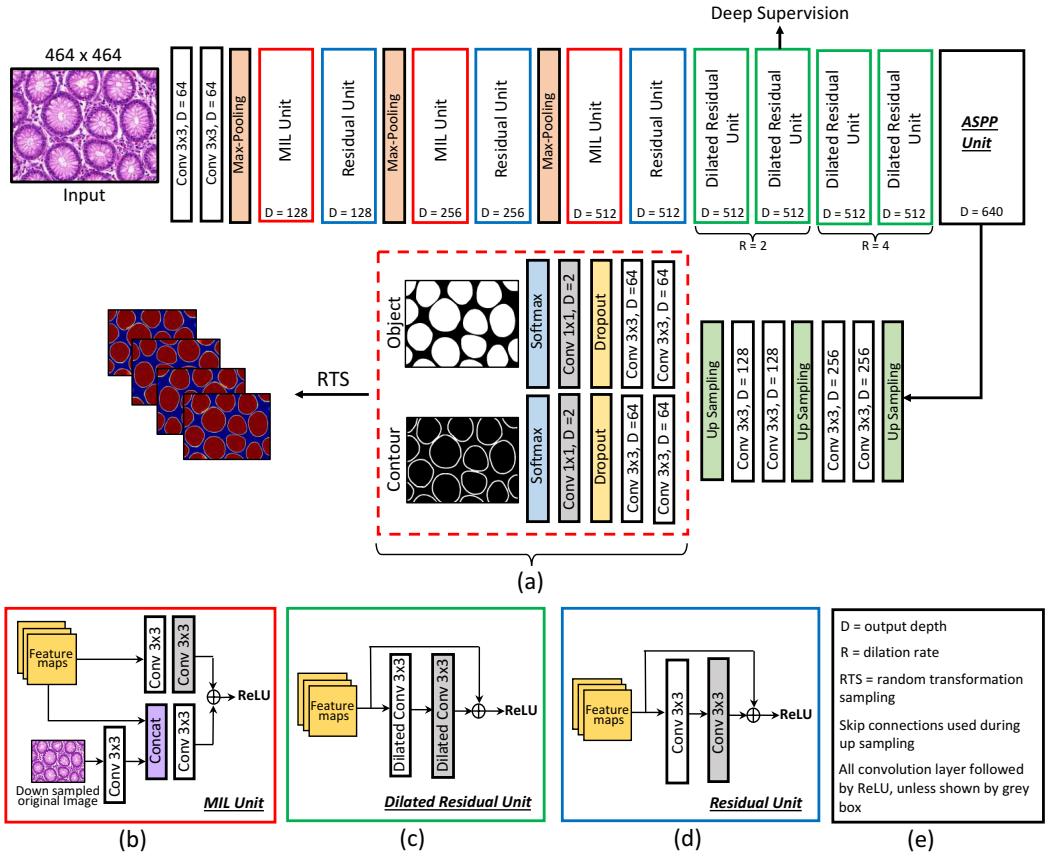


Figure 4.2: Overview of the proposed network architecture for gland instance segmentation.

ure 4.2. When we use a downsampled version of the original image (downsampled with bicubic interpolation) without max-pooling, it indirectly captures the variation in pixel intensities in the local neighbourhood of each pixel without completely discarding the activations, as is the case with max-pooling. It is this principle that allows the MIL unit to ensure that the missing details are preserved. Equation 4.1 is modified to generate the MIL unit. The MIL unit can be defined as:

$$\boldsymbol{y} = \mathcal{F}(\boldsymbol{x}, \boldsymbol{W}) + \mathcal{G}(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{M}) \tag{4.2}$$

where $\mathcal{F}$ is defined in the same way as Equation 4.1. The vector $\boldsymbol{v}$ denotes the original downsampled image and is incorporated into the function $\mathcal{G}$ to minimise the loss of information. $\mathcal{G}$ represents the function $\boldsymbol{M}_2(\sigma(\boldsymbol{M}_1\boldsymbol{v})\|\boldsymbol{x})$, where $\|$ denotes the concatenation operation. Similar to the traditional residual unit, $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ within function $\mathcal{G}$ represent the weights of the convolution applied to the downsampled image and the convolution of the concatenated feature maps respectively. The

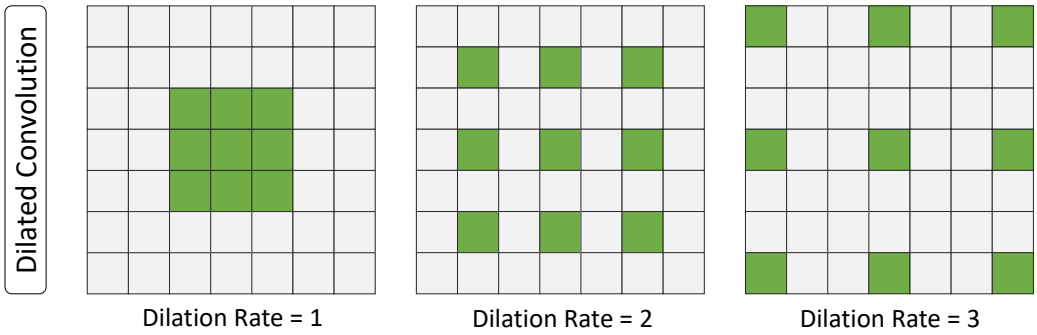


Figure 4.3: Illustration of dilated convolution with varying dilation rates. The green squares denote the position of a 3×3 filter acting on an image/feature map. Using a dilation rate of 1 is the same as a regular convolution. Increasing the dilation rate introduces sparsity in the kernel and increases the size of the receptive field.

summation of $\mathcal{F}$ and $\mathcal{G}$ is shown by the $\oplus$ symbol in the MIL unit within Figure 4.2.

Instead of downsampling the size of the input to increase the size of the receptive field, an alternate solution is to increase the size of the kernel during convolution. However, this practice is not feasible due to the huge amount of parameters required. Instead, dilated convolution uses sparse kernels [165], such that the resolution of the feature maps is preserved, without significantly increasing the number of parameters. In Figure 4.3 we display an illustration of dilated convolution with a 3×3 kernel and different dilation rates. We can see that using a dilation rate of 1 is the same as a regular convolution, whereas using a dilation rate greater than 1 introduces sparsity within the kernel and consequently increases the size of the receptive field. We incorporate dilated convolution into residual units simply by replacing each 3×3 convolution with a 3×3 dilated convolution. We initially downsample using max-pooling and MIL units and then use dilated convolution when the image has been downsampled by a factor of 8. We do not use dilated convolution throughout the entire network since otherwise the model does not fit into GPU memory. This is because convolving over the size of the original image is more computationally expensive compared to when this image is downsampled. Dilated residual units can be seen in part (b) of Figure 4.2. Minimising the loss of information allows us to perform a successful gland instance segmentation, without the need to incorporate additional information that is used in other methods [26]. Retaining the information throughout the model allows the network to successfully segment small glandular objects and thin glandular contours. It must be noted that we output the contours for uncertainty map refinement; not for separating gland

instances.

In addition, for effective multi-scale aggregation, we apply *atrous* spatial pyramid pooling (ASPP) [28] to the output of the deep network. Within our framework, the goal of ASPP is to combat the challenge of detecting glands of different cancer grades that display a high level of morphological heterogeneity. To achieve this, we merge together multiple dilated convolution layers, allowing us to explicitly control the size of the receptive field. Specifically, we use three dilated convolution operations, with rates 6, 12 and 18. When the dilation rate is large, the dilated convolution reduces to a 1×1 convolution. This is because the dilated kernel becomes larger than the size of the input feature map. Instead, to incorporate global level context, we also use global average pooling. All operations are followed by an initial 1×1 convolution, a dropout layer with a rate of 0.5 and then a second 1×1 convolution for reducing the depth of the output. The concatenation of these feature maps gives a powerful representation of the features extracted from the minimal information loss dilated network.

Although high-level contextual information can be generated within the deep neural network, it is crucial to incorporate low-level information for precisely delineating the glandular boundaries. Directly upsampling by a factor of 8 to produce the output does not consider low-level information. Instead, similar to U-Net [121], we choose to upsample by a factor of 2 each time and concatenate low-level features to the start of each upsampling block. Before the concatenation, we apply a 1×1 convolution to increase the depth of lower levels; ensuring that we have an equal contribution of both components during the concatenation. We concatenate the feature maps from the second convolution layer and the first two standard residual units. We find that this method of upsampling is especially important for precisely locating the boundaries where low-level features are particularly important. When the features have been upsampled to the resolution of the original image, the network splits into two separate branches: one for the gland object and one for the contour. We denote this part of the network the task specific component of the network and is shown by the dashed red box in Figure 4.2(a). We show an example of how the task specific component can be modified in Section 2.3 of this chapter. We add deep supervision to our network by calculating the auxiliary loss at the second dilated residual unit during feature extraction. This helps the network to learn more discriminative features and encourages a faster convergence. We also add dropout layers immediately before the final 1×1 convolution, near the output of the network, with a rate of 0.5. The overall flow of the network can be seen in Figure 4.2.

4.2.2 MILD-Net Loss Function

During training, we calculate the cross-entropy loss with respect to all outputs of the proposed network. Concretely, we define $\mathcal{L}_g$, $\mathcal{L}_c$, $\mathcal{L}_{a_g}$, $\mathcal{L}_{a_c}$ to be the gland, contour, gland auxiliary and contour auxiliary cross-entropy loss functions respectively. We define each loss function as:

$$\begin{aligned}
\mathcal{L}_g &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\Psi_{i,k}(\boldsymbol{x})\log p_i(\boldsymbol{x}, w_g) \\
\mathcal{L}_c &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\Phi_{i,k}(\boldsymbol{x})\log q_i(\boldsymbol{x}, w_c) \\
\mathcal{L}_{a_g} &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\Psi_{i,k}(\boldsymbol{x})\log r_i(\boldsymbol{x}, w_{a_g}) \\
\mathcal{L}_{a_c} &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\Phi_{i,k}(\boldsymbol{x})\log s_i(\boldsymbol{x}, w_{a_c})
\end{aligned} \tag{4.3}$$

where $p_i(\boldsymbol{x}, w_g)$, $q_i(\boldsymbol{x}, w_g)$, $r_i(\boldsymbol{x}, w_{a_g})$ and $s_i(\boldsymbol{x}, w_{a_c})$ represent the softmax output at pixel i for the gland, contour, auxiliary gland and auxiliary contour outputs. Then, Ψ is the gland ground truth, Φ is the contour ground truth and K is the number of classes which we set to 2 because we perform binary segmentation at each output. To calculate the overall loss at the output of each branch, we average the cross-entropy over all N pixels. Then, the final loss function to be minimised during training is defined as:

$$\mathcal{L} = \mathcal{L}_g + \mathcal{L}_c + \lambda\mathcal{L}_{a_g} + \lambda\mathcal{L}_{a_c} + \gamma||\boldsymbol{w}||_2^2 \tag{4.4}$$

where discount weight λ decays the contribution of each auxiliary loss $\mathcal{L}_{a_g}$ and $\mathcal{L}_{a_c}$ during training. We initially set λ as 1, and divide the value by 10 after every eight training epochs. The selection of the initial λ and the decay strategy was motivated by DCAN [26], where they used a similar strategy. $||\boldsymbol{w}||_2^2$ denotes the regularisation term on weights $\boldsymbol{w} = \{\boldsymbol{w}_g, \boldsymbol{w}_c, \boldsymbol{w}_{a_g}, \boldsymbol{w}_{a_c}\}$, with regularisation parameter γ. We emperically set gamma to be 10^{-5}.

4.2.3 Random Transformation Sampling for Uncertainty Quantification

Current deep learning models have an ability to learn powerful feature representations and are capable of successfully mapping high dimensional input data to an output. However, this mapping is assumed to be accurate in such models and there is no quantification of how certain the model is of the prediction. Bayesian approaches to modeling, naturally involve uncertainty quantification by obtaining a posterior distribution over the parameters of the model, which therefore allows us to induce a predictive distribution for the unseen data. However, the tractability and scalability of Bayesian methods applied to shallow neural networks and their recent deeper counterparts have been a subject of research for the past several decades. Although significant progress has been made, inference of the posterior distribution over the model parameters remains computationally expensive. Recent work [51] demonstrated that using a standard regularisation tool such as dropout is equivalent to variational approximation using Bernoulli distributions [22] in deep learning. Therefore, this can be used to approximate the uncertainty over the model predictions [50]. Standard variational dropout captures the uncertainty over the model weights, given the observed data. It is important to distinguish that there may be noise inherent to each observation, that we might not be able to reduce by obtaining more data. This would be crucial to estimate within clinical applications. Generally, this uncertainty is estimated through a data dependent noise model [80], however it would require us to modify the existing architecture. Therefore, to capture observation dependent noise, we perform random transformations to the input images during test time. To obtain the predictive distribution, we apply a random transformation $\Phi(\mathbf{x})$ on a sample of n images, where Φ performs a flip, rotation, Gaussian blur, median blur or adds Gaussian noise on input image $\mathbf{x}$ to obtain $\{\Phi_1, \Phi_2, ..., \Phi_n\}$. Each image within the sample is then processed, where the mean of this processed sample gives the refined prediction and the variance gives the uncertainty. Due to the aggregation of the predictions of multiple transformed images, our model will naturally perform well, particularly for areas that are generally difficult to classify. Similarly, recent work leveraged transformed images, but instead are utilised to obtain informative priors [112], that help a model become more invariant to these specific transformations. However, the primary aim for utilising RTS is to obtain a measure of uncertainty that may be informative within clinical practice, as opposed to making our model more invariant. We can define the prediction and uncertainty as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} f(\Phi_i(\boldsymbol{x}); \boldsymbol{w}); \quad \sigma = \frac{1}{N}\sum_{i=1}^{N} \left(f(\Phi_i(\boldsymbol{x}); \boldsymbol{w}) - \mu\right)^2 \tag{4.5}$$

where μ defines the segmentation prediction, σ defines the uncertainty and N defines the number of transformations. The function f denotes the deep neural network with input $\boldsymbol{x}$ and output taken after the softmax layer. $\boldsymbol{w}$ denotes the weights and Φ_i defines a random transformation i to input image $\boldsymbol{x}$. Note, that the output of σ is a two-dimensional image, where high values denote pixels with high uncertainty.

We propose a metric to give individual glands a score of uncertainty, based on the uncertainty map generated via random transformation sampling (RTS). This measure highlights glands that are generally hard to classify, irrespective of the number training examples that the model has seen. We suggest that it is reasonable to disregard segmented glands that have an uncertainty score above a given threshold, because in practice features would not be extracted from areas of general ambiguity. We first remove the boundaries by subtracting the predicted contours that have been output by the network and then calculate the object-level uncertainty score for each predicted instance k as: $\tau_k = \frac{1}{N}\sum_{i=1}^{N} \hat{\sigma}\rho_{k,i}$, where $\hat{\sigma}$ is the boundary removed uncertainty map and $\rho_{k,i}$ is the predicted binary output of pixel i within instance k. We define n as the number of pixels within predicted instance k. We remove the boundaries because these areas show the transition between the two classes and therefore the uncertainty here can't be avoided. Given a selected global threshold for our uncertainty score τ, we may only consider segmented glands with a score below this threshold.

4.2.4 MILD-Net$^+$ for Simultaneous Gland and Lumen Segmentation

We extend MILD-Net such that it simultaneously segments the lumen and the gland, in order to increase the diagnostic power of the network. For example, when the grade of cancer increases, tumours tend to become solid and lose their lumenal properties. Therefore, the additional segmentation of the lumen can empower current automated colorectal cancer classification methods, due to the introduction of additional important diagnostic features. In order to achieve this simultaneous segmentation, the network requires only a subtle modification. MILD-Net$^+$ takes an image as input and, identically to MILD-Net, extracts features via the minimal information loss encoder. After upsampling to the original resolution, the task-specific component of the network is modified such that it has four branches. The only difference between MILD-Net and MILD-Net$^+$ is the number of branches after the

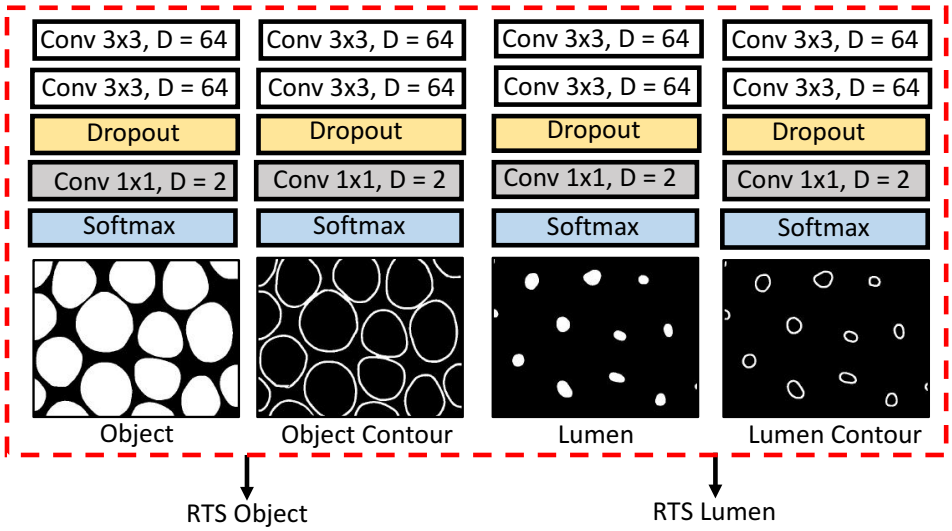


Figure 4.4: Modification of network output for simultaneous gland and lumen segmentation.

network is upsampled back to the size of the original image. Specifically, the part of the architecture shown by the red dashed box displayed in Figure 4.2(a) is replaced with the one in Figure 4.4. We observe that the majority of the network is unchanged apart from the addition of two branches at the end of the upsampling path. As a result, MILD-Net$^+$ does not require many additional parameters to achieve an accurate and simultaneous gland and lumen segmentation. This highlights the ability of MILD-Net$^+$ to extract a rich set of features. Similar to what we have done before, we apply RTS to both the gland and the lumen and use the gland and lumen contours to refine the output of each uncertainty map. Consequently, MILD-Net$^+$ segments diagnostically important features, whilst quantifying the uncertainty for each segmented component.

4.2.5 MILD-Net$^+$ Loss Function

In the same fashion as Section 2.2, we calculate the cross-entropy loss with respect to the output of each component of MILD-Net$^+$. Specifically, we calculate the cross entropy loss with respect to the gland, gland contour, lumen and lumen contour denoted by $\mathcal{L}_g$, $\mathcal{L}_{gc}$, $\mathcal{L}_l$ and $\mathcal{L}_{lc}$ respectively. We aso calculate the auxiliary losses $\mathcal{L}_{a_g}$ and $\mathcal{L}_{a_l}$ with respect to the gland and the lumen. Then, during training, the overall loss function of MILD-Net$^+$ is defined as:

$$\mathcal{L} = \mathcal{L}_g + \mathcal{L}_{gc} + \mathcal{L}_l + \mathcal{L}_{lc} + \lambda\mathcal{L}_{a_g} + \lambda\mathcal{L}_{a_l} + \gamma||\boldsymbol{\theta}||_2^2 \tag{4.6}$$

where $||\boldsymbol{\theta}||_2^2$ denotes the regularisation term on weights $\boldsymbol{\theta} = \{\boldsymbol{\theta}_g, \boldsymbol{\theta}_{gc}, \boldsymbol{\theta}_l, \boldsymbol{\theta}_{lc}, \boldsymbol{\theta}_{a_g}, \boldsymbol{\theta}_{a_l}\}$, with regularisation parameter γ. We use the same γ as MILD-Net, with a value of 10^{-5}. Also, we use the same λ that was utilised within MILD-Net that decays the contribution of the auxiliary loss during training. In a similar vein, we also divide the value by 10 after every eight training epochs. Note, that we choose not to use auxiliary loss with respect to the contours in order to reduce the number of parameters in MILD-Net$^+$.

4.3 Experiments and Results

4.3.1 The Datasets and Pre-processing

For our experiments, we used two datasets: (i) the Gland Segmentation (GlaS) challenge dataset [135], used as part of MICCAI 2015, and (ii) a second independent colon adenocarcinoma dataset, which for simplicity we refer to as the colorectal adenocarcinoma gland (CRAG) dataset[1], that was originally used in [15]. Both datasets were obtained from the University Hospitals Coventry and Warwickshire (UHCW) NHS Trust in Coventry, United Kingdom. Within (i), data was extracted from 16 H&E stained histological WSIs, scanned with a Zeiss MIRAX MIDI Slide Scanner with a pixel resolution of 0.465μm/pixel. After scanning, the WSIs were rescaled to 0.620μm/pixel (equivalent to 20$\times$ objective magnification) and then a total of 165 image tiles were extracted. These 165 images consist of 85 training (37 benign and 48 malignant) and 80 test images (37 benign and 43 malignant). Furthermore, the test images are split into two test sets: Test A and Test B. Test A was released to the participants of the GlaS challenge one month before the submission deadline, whereas Test B was released on the final day of the challenge. Further information on the dataset can be found within the published challenge paper[135]. Images are mostly of size 775$\times$522 pixels and all training images have associated instance-level segmentation ground truth that precisely highlight the gland boundaries. In addition, two expert pathologists (D.S, Y.W.T) provide accurate lumen annotations for all glands within the GlaS dataset. Within (ii), we have a total of 213 H&E CRA images taken from 38 WSIs scanned with an Omnyx VL120 scanner with a pixel resolution of 0.55μm/pixel (20$\times$ objective magnification). All 38 WSIs are from different patients and are mostly of size 1512$\times$1516 pixels, with corresponding instance-level ground truth. The CRAG dataset is split into 173 training images and 40 test images with different cancer grades. For both datasets, we set 20% of

[1]The CRAG dataset for gland segmentation is available at `https://warwick.ac.uk/fac/sci/dcs/research/tia/data/mildnet`

the training set aside for evaluating the performance of our model during training. Examples of images from each of the two datasets can be seen in Figure 4.1.

We extracted patches of size 500×500 and augmented patches with elastic distortion, random flip, random rotation, Gaussian blur, median blur and colour distortion. Finally, we randomly cropped a patch of size 464×464, before input into the proposed network.

4.3.2 Whole-Slide Image Processing

In addition to processing the image tiles as described in Section 3.1 of this chapter, we further investigated the ability of our method by processing a set of colorectal adenocarcinoma WSIs. This dataset consists of 16 high resolution WSIs, taken from the COMET dataset, which was originally used in [134]. Within this dataset, the WSIs are obtained from two different centres and therefore we split the images into two further datasets. We name the dataset corresponding to WSIs from the first centre as COMET-1 and the dataset containing WSIs from the second centre as COMET-2. COMET-1 is from the same centre as the image tiles that the algorithm was trained on, whereas COMET-2 is from a different centre completely. We introduce the second centre to test how our method generalises to new data. The data is divided equally, such that 8 WSIs are taken from each centre. Because it is quite laborious to obtain pixel-based glandular annotations for each WSI, we select two high-power fields (HPFs) from each WSI of size 2500×2500 pixels at 20×. As a result, even though we process the whole-slide to see how our algorithm performs visually, we use these selected HPFs to perform quantitative comparison. HPFs were extracted such that we had an even representation of benign and malignant regions, annotated by two expert pathologists (D.S, Y.W.T). In order to satisfy this criteria, we mainly processed WSIs that contained a combination of malignant and benign glands.

4.3.3 Implementation and Training Details

We implemented our framework with the open-source software library TensorFlow version 1.3.0 [7]. We used Xavier initialisation [55] for the weights of the model, where they were drawn from a Gaussian distribution. Concretely, weight w_i is initialised with mean 0 and variance $\frac{1}{n_{w_i}}$, where, n_{w_i} refers to the number of input neurons to weight i. We trained our model on a workstation equipped with one NVIDIA GeForce Titan X GPU for 30 epochs (60,000 steps) on the GlaS dataset and 75 epochs (200,000 steps) on the CRAG dataset. The difference in the number

of steps until convergence reflects the greater variability of the CRAG dataset. We used Adam optimisation with an initial learning rate of 10^{-4} and a batch size of 2.

4.3.4 Evaluation and Comparison

We assessed the performance of our method by using the same evaluation criteria used in the MICCAI GlaS challenge, consisting of F_1 score, object-level dice and object-level Hausdorff distance [135]. The F1 score is employed to measure the detection accuracy of individual glandular objects, the Dice index is a measure of similarity between two sets of samples and the Hausdorff distance measures the boundary-based segmentation accuracy. We implemented several state-of-the-art segmentation methods including SegNet [17], FCN-8 [103] and a DeepLab-v3 [28] model for extensive comparative analysis. For gland segmentation, we also report the results obtained by two recent methods including MIMO-Net [118], that uses a multi-input-multi-output convolutional neural network and two methods that utilise deep multichannel side supervision [160, 161].

For all methods, including MILD-Net, the final binary maps are obtained by applying a threshold of 0.5 to all predicted probability maps. A morphological opening operation is then used with a disk filter radius 5 to obtain the final result. This disk size was empirically selected because it gave the best visual and quantitative results.

In this section, we first show results for MILD-Net on the GlaS dataset and the CRAG dataset. Next, we display results of MILD-Net for whole-slide image (WSI) processing. Finally, we report results of MILD-Net$^+$ on the GlaS dataset.

Results on GlaS and CRAG Datasets Using MILD-Net

We can see from Table 4.1 that our proposed network achieves state-of-the-art performance compared to all methods on the 2015 MICCAI GlaS Challenge dataset. We also validated the efficacy of our method on the CRAG dataset, demonstrating overall better performance in comparison with other methods and highlighting the good generalisation capability of our method on different datasets. Results on the CRAG dataset can be seen in Table 4.2. We can see from Table 4.3 that utilising test time random transformations leads to an improved performance, due to a refined prediction within areas of high uncertainty. Additionally, we compared our method of RTS to Monte Carlo dropout sampling. However, because we don't apply many dropout layers within our network, there is not sufficient variation in the samples to have a profound effect. We also experimented by adding additional dropout

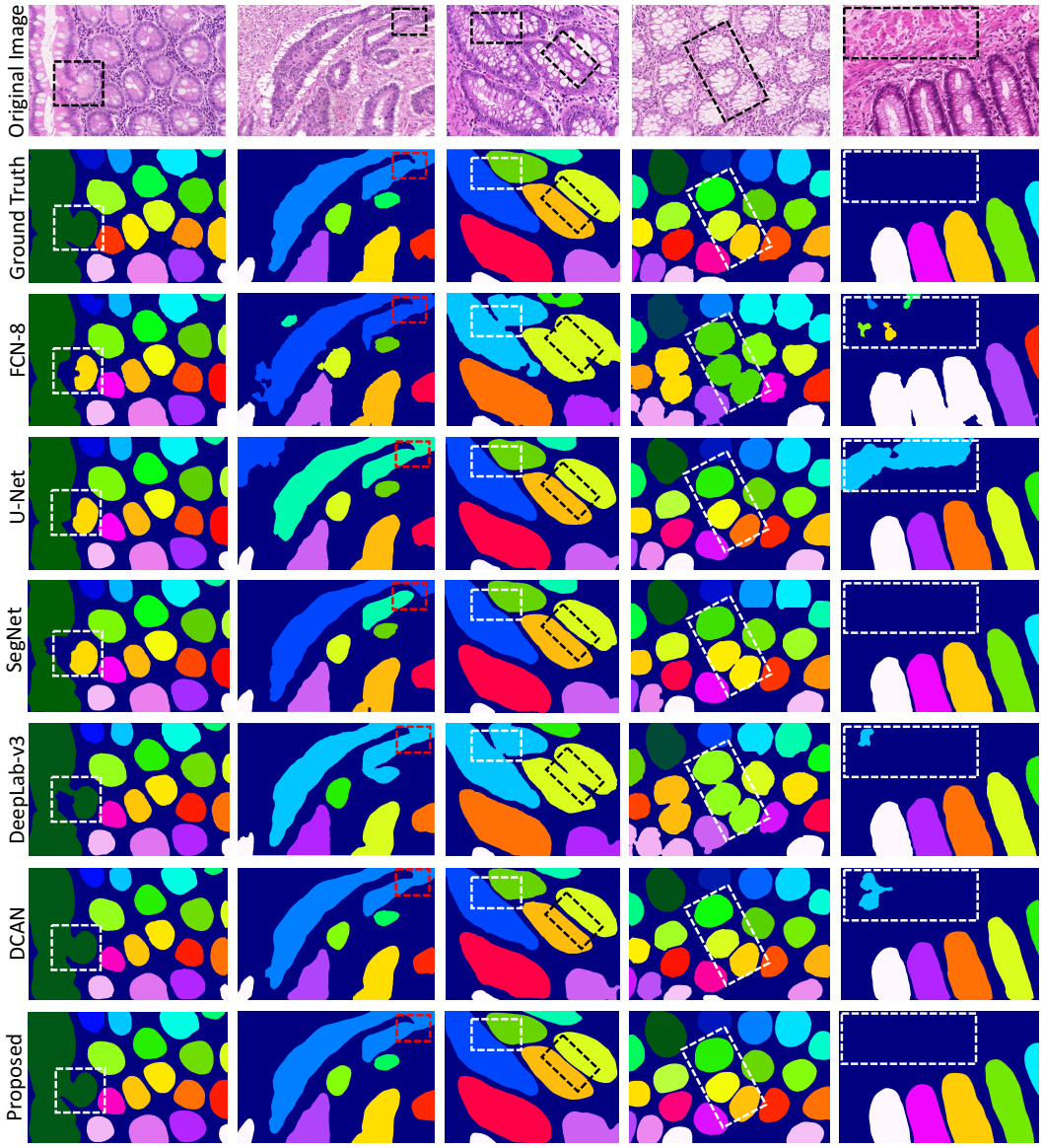


Figure 4.5: Visual gland segmentation results on the GlaS dataset.

layers with Monte Carlo dropout, but this had a detrimental effect during the training of the network. Because RTS utilises an averaging technique, the number of false positives in areas of high uncertainty is reduced. This explains the increase in performance with RTS. It must be noted that it is significantly more difficult to segment glands within the CRAG dataset than when using the GlaS dataset. This is because there are many malignant cases where the glandular boundaries are very ambiguous. Examples of results from different methods are shown in Figure 4.5 and 4.6. We can see that our method can generate more accurate gland instance

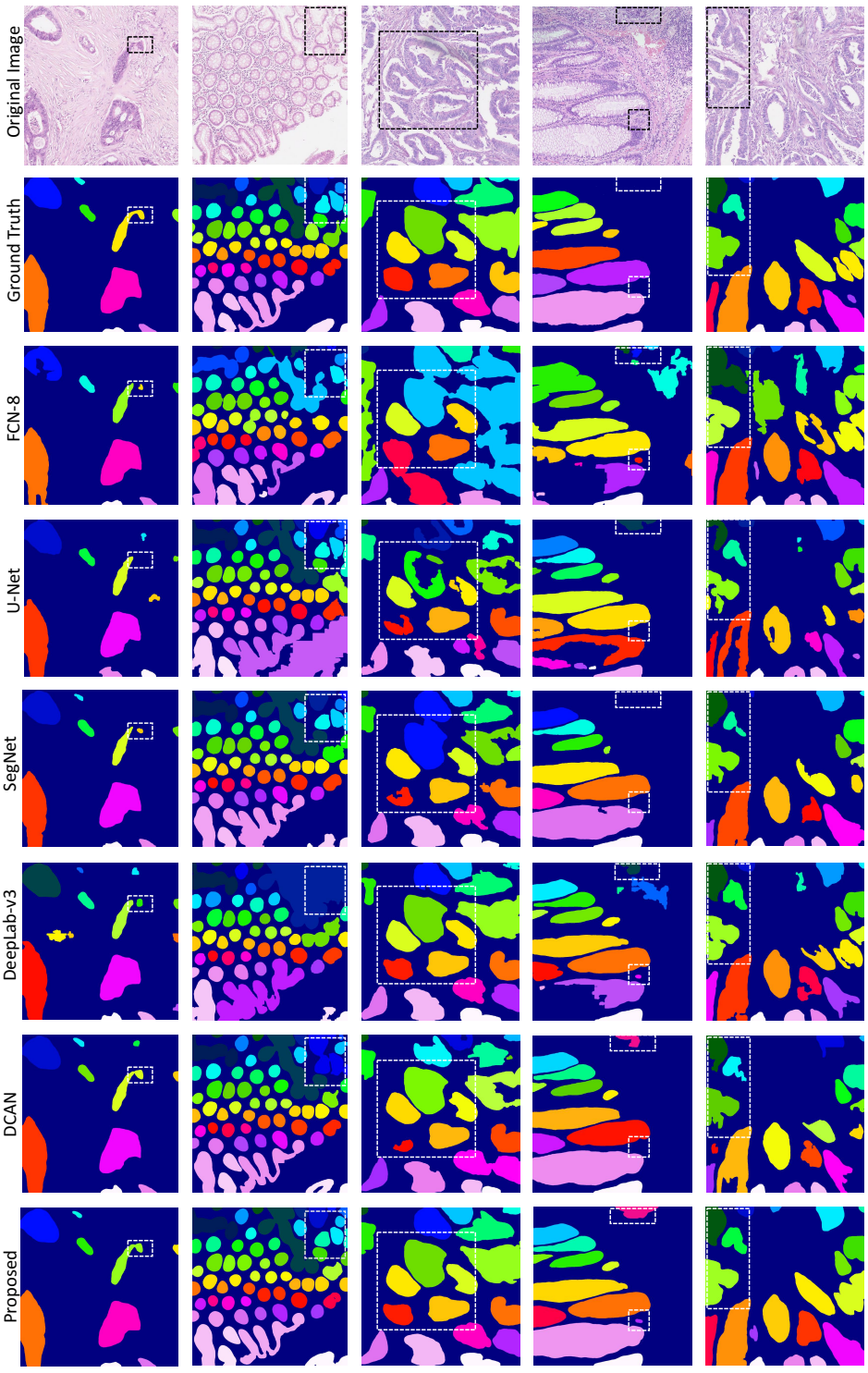


Figure 4.6: Visual gland segmentation results on the CRAG dataset.

Table 4.1: Comparative analysis of models on the GlaS challenge dataset. CUMedVision submissions use the method reported in [27] and Freidburg submissions use the method reported in [121].

Method	F_1 Score		Obj. Dice		Obj. Hausdorff		Rank Sum
	Test A	Test B	Test A	Test B	Test A	Test B	
MILD-Net	**0.914**	**0.844**	**0.913**	**0.836**	**41.54**	**105.89**	**6**
Multichannel B [161]	0.893	0.843	0.908	0.833	44.13	116.82	15
MIMO-Net [118]	0.913	0.724	0.906	0.785	49.15	133.98	31
Multichannel A [160]	0.858	0.771	0.888	0.815	54.20	129.93	33
DeepLab [28]	0.862	0.764	0.859	0.804	65.72	124.97	46
SegNet [17]	0.858	0.753	0.864	0.807	62.62	118.51	46
FCN-8 [103]	0.783	0.692	0.795	0.767	105.04	147.28	71
CUMedVision2 [27]	0.912	0.716	0.897	0.781	45.42	160.35	43
ExB1	0.891	0.703	0.882	0.786	57.41	145.58	49
ExB3	0.896	0.719	0.886	0.765	57.36	159.87	52
Freidburg2 [121]	0.87	0.695	0.876	0.786	57.09	148.47	52
CUMedVision1 [27]	0.868	0.769	0.867	0.8	74.6	153.65	54
ExB2	0.892	0.686	0.884	0.754	54.79	187.44	61
Freidburg1 [121]	0.834	0.605	0.875	0.783	57.19	146.61	63
CVML	0.652	0.541	0.644	0.654	155.43	176.24	94
LIB	0.777	0.306	0.781	0.617	112.71	190.45	95
vision4GlaS	0.635	0.527	0.737	0.61	107.49	210.1	98

Table 4.2: Comparative analysis of models on the CRAG dataset. S and R denote score and rank respectively.

Method	F_1 Score	Obj. Dice	Obj. Hausdorff	Rank Sum
MILD-Net	**0.825**	**0.875**	**160.14**	**3**
DCAN [27]	0.736	0.794	218.76	6
DeepLab [28]	0.648	0.745	281.45	10
SegNet [17]	0.622	0.739	247.84	11
U-Net [121]	0.600	0.654	354.09	15
FCN-8 [103]	0.558	0.640	436.43	18

Table 4.3: MILD-Net performance with random transformation sampling (RTS) on the CRAG and GlaS datasets.

Dataset	Method	F_1 Score	Obj. Dice	Obj. Hausdorff
GlaS A	MILD-Net	0.914	0.908	42.32
	MILD-Net-RTS	0.914	**0.913**	**41.54**
GlaS B	MILD-Net	0.809	0.822	117.91
	MILD-Net-RTS	**0.844**	**0.836**	**105.89**
CRAG	MILD-Net	0.806	0.867	162.35
	MILD-Net-RTS	**0.825**	**0.875**	**160.14**

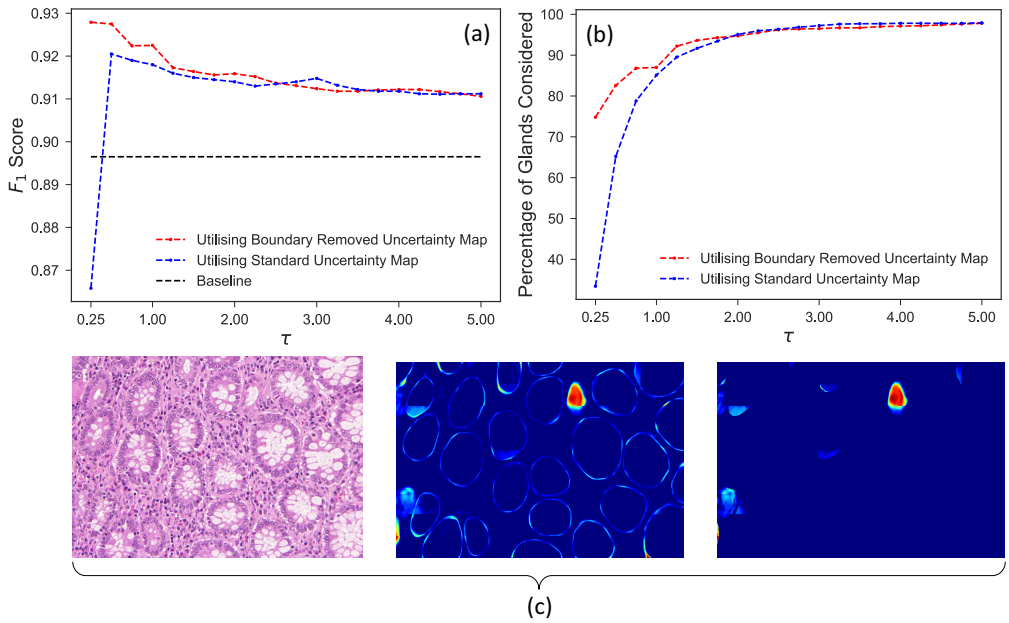


Figure 4.7: Object-level uncertainty quantification. (a) F_1 score as we disregard predictions with an uncertainty score τ_k greater than a given threshold τ. (b) Percentage of total instances considered, given a threshold τ. (c) From left to right: original image; uncertainty map σ; boundary removed uncertainty map $\hat{\sigma}$.

segmentation with precisely delineated boundaries and well segmented instances. It is interesting to see that within the dashed boxes in the last column of Figure 4.6, our proposed algorithm was able to detect tumorous areas that were not picked up by the pathologist.

In Figure 4.7, we show the relationship between the performance and the uncertainty score τ. This score is used as a threshold, where we only consider predictions k with an uncertainty score τ_k lower than τ. We observe from Figure 4.7 that it seems sensible to only consider segmented predictions with an uncertainty score τ_k below 1. This preserves a large proportion of the dataset, whilst significantly increasing the performance. We also display the effect of using the boundary removed uncertainty map. We observe that removing the boundary allows us to preserve a larger proportion of the dataset when we are using lower thresholds for the removal of predictions with high uncertainty. This suggests that using the boundary removed uncertainty map allows us to correctly remove the uncertain cases that contribute most negatively to the performance. Therefore, utilising the boundary removed uncertainty map is more robust and can be effectively be used to select predictions with low uncertainty. It is interesting to note that we are still able to preserve around 75% of instances by selecting predictions with τ_k below 0.25. As a

result, F_1 score, object dice and object Hausdorff can be increased to 0.930, 0.9359 and 28.658 for test set A and increased to 0.913, 0.9567 and 22.70 for test set B. It must be noted that the intuition of disregarding glands with high uncertainty means that we should not extract any statistical measures from these disregarded glands. Therefore, when removing predicted instances with high uncertainty, we also remove the corresponding ground truth instance to obtain the above measures.

Results on Whole-Slide Images Using MILD-Net

Within part (a-d) of Figure 4.8, the inner-most image is the original WSI with overlaid glandular boundaries, the central column shows the two HPFs for statistical analysis at 20× and the outer-most column shows a selected region of each HPF at 40×. We observe that our proposed method is able to accurately segment glands within colon whole-slide histology images with a precise delineation of glandular boundaries. Therefore, as a result of training on both the GlaS and the CRAG dataset, our method is capable of extracting a strong set of features that enables a successful transition to WSI processing. We also note from part (c) and (d) of Figure 4.8, that MILD-Net generalises well to completely unseen data from different centres. A particularly interesting aspect of COMET-2 is that most images contain pathologist pen markings. However, as a result of the strong set of features that MILD-Net is able to extract no pre-processing was needed to avoid these regions, where other methods may have failed. For a thorough analysis, we obtain quantitative results for all HPFs extracted from the 16 WSIs. In total, we have 32 HPFs: 16 from COMET-1 and 16 from COMET-2. In order to test the performance of our algorithm on both benign and malignant cases, we ensured an equal representation of both benign and malignant glands. We can see from Table 4.4 that the proposed method has a similar performance between the two datasets, highlighting the generalisability of our method. Despite a good detection performance, we can see that the Hausdorff distance within malignant cases is significantly higher than those results reported on the GlaS and the CRAG dataset. The Hausdorff distance measure indicates how closely the shape of two objects match with each other. As a result, disagreement at the boundary will lead to deterioration in performance. Therefore, this suggests that the algorithm finds it challenging to precisely locate the glandular boundaries within malignant cases. This however reflects the true difficulty in segmenting glands within whole-slide histology images, where there are often many ambiguous regions. After careful observation, we state that the lower performance for Hausdorff distance is not due to a limitation of the algorithm, but because a number of malignant cases are generally difficult to segment.

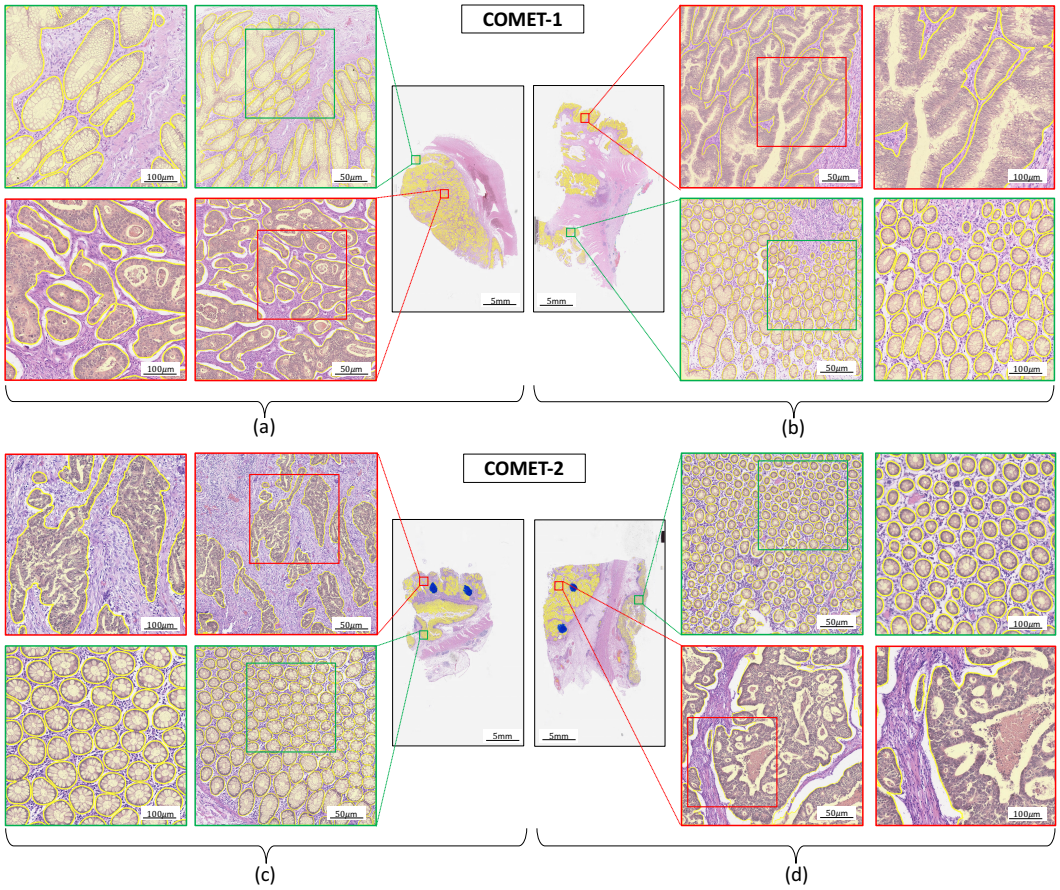


Figure 4.8: Visual results of gland segmentation on WSIs using the proposed framework.

Table 4.4: MILD-Net gland segmentation performance on HPFs from WSIs. B stands for average benign score and M stands for average malignant score.

	F_1 Score		Obj. Dice		Obj. Hausdorff	
	B	**M**	**B**	**M**	**B**	**M**
COMET-1	0.811	0.817	0.822	0.867	158.40	389.89
COMET-2	0.948	0.716	0.886	0.751	76.15	474.12
Average COMET-1	0.814		0.845		274.15	
Average COMET-2	0.832		0.819		275.14	

Results on GlaS Dataset Using MILD-Net$^+$

To demonstrate the performance of MILD-Net$^+$, we compare our algorithm to four recent segmentation methods trained solely for the task of lumen segmentation. Namely, these methods are FCN-8 [103], U-Net [121], SegNet [17] and DeepLab-v3 [28]. We chose not to compare with DCAN [27] because this network was specifi-

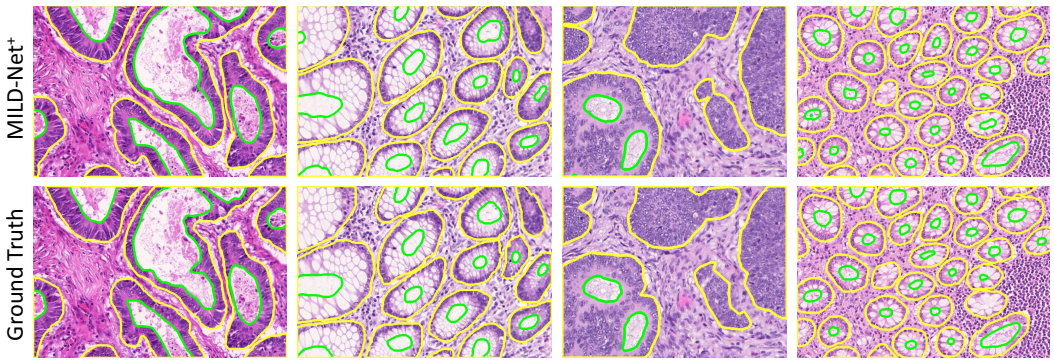


Figure 4.9: Visual results for simultaneous gland and lumen segmentation.

cally tuned to achieve instance segmentation. Instance segmentation is not an issue for lumen segmentation, because neighbouring lumen physically can't touch within histology images. The only exception for this would be if there was an artefact within the image. From Figure 4.9, we observe that our algorithm is able to precisely segment both the gland object and the gland lumen. We can see in Table 4.5, that MILD-Net$^+$ achieves superior performance in all statistical measures for lumen segmentation, compared to all competing methods. This is particularly interesting because all other competing methods were trained for the single task of lumen segmentation. Therefore, this reiterates the strong feature extraction capabilities of the minimal information loss network. Despite achieving state-of-the-art performance at the output of the lumen branch, it is necessary to ensure that we still achieve a good accuracy at the output of the gland object branch. We observe that, MILD-Net$^+$ out-performs MILD-Net on most of the statistical measures, suggesting that segmenting the lumen may provide additional cues to strengthen the segmentation of the gland object.

Table 4.5: MILD-Net$^+$ gland and lumen segmentation performance on the GlaS challenge dataset.

		F_1 **Score**		**Obj. Dice**		**Obj. Hausdorff**	
		Test A	**Test B**	**Test A**	**Test B**	**Test A**	**Test B**
Lumen	**MILD-Net$^+$**	**0.825**	**0.711**	**0.875**	**0.816**	**26.81**	**94.09**
	DeepLab [28]	0.757	0.521	0.816	0.722	46.49	136.81
	SegNet [17]	0.698	0.661	0.791	0.781	56.22	110.32
	U-Net [121]	0.623	0.425	0.724	0.643	73.51	152.52
	FCN-8 [103]	0.744	0.556	0.778	0.723	60.51	133.09
Gland	**MILD-Net$^+$**	**0.920**	0.820	**0.918**	**0.836**	**39.39**	**103.07**
	MILD-Net	0.914	**0.844**	0.913	**0.836**	41.54	105.89
	CUMedVision2 [27]	0.912	0.716	0.897	0.781	45.42	160.35

4.4 Discussion and Conclusions

Analysis of Hematoxylin and Eosin stained histology slides is considered as the *gold standard* in histology based diagnosis. However, visual analysis is very time consuming and laborious because pathologists are required to thoroughly examine each case to ensure an accurate diagnosis. Furthermore, due to the complex nature of the task, histopathological diagnosis often suffers from inter- and intra-observer variability. Computational techniques aim to counter the challenges posed within routine pathology by providing an objective and potentially more accurate diagnosis. In order to improve the diagnostic capabilities of automated methods, we present a minimal information loss dilated network for the accurate segmentation of glands within colon histology images. Subsequently, gland based features can be used to empower the diagnostic decision made by the pathologist.

Extensive experimentation on multiple datasets demonstrates the superior performance of our approach compared to other competing methods. Furthermore, our method performs well when applied to the WSI, highlighting the network's strong feature extraction capabilities. As a result, the network may be used in a clinical setting to segment glandular structures within the WSI with a high level of accuracy. We also show that the method generalises well to new data and can therefore be expected to work well within other centres.

It is worth noting, that the minimal information loss network helps retain the spatial information within the network and therefore leads to a successful segmentation at the glandular boundaries. Therefore, additional cues are not needed to separate the majority of touching instances. However, it must be noted that this method is able to separate glands when they are very close together, but may fail when the glands are physically touching with no pixels in between. We do not see this as a cause for concern because the majority of instances can be separated by our method due to the reduction of information loss throughout the network. We also observe from our results that our network was able to successfully segment glands of various sizes. This in part was because of the addition of the *atrous* spatial pyramid pooling module that enlarged the size of the receptive field with varying dilation rates.

The addition of RTS increased the performance of the algorithm, whilst simultaneously generating an uncertainty map. We have shown that this uncertainty map can be used as additional information about where the algorithm is uncertain. Also, we have shown that if we choose not to extract features from predictions with high uncertainty, we can significantly increase the performance whilst maintaining

a large proportion of the dataset. We can ensure that we retain a larger proportion of this dataset if we use a boundary removed uncertainty map. The removal of predictions with high uncertainty is particularly important for gland-based feature extraction (e.g glandular aberrance [15]) because features should not be extracted from glands where the algorithm is not confident. This workflow mimics clinical practice because the pathologist would not make a diagnosis from areas of ambiguity. Therefore, this uncertainty map can be used to extract relatively strong features for subsequent grading.

The proposed network may fail to distinguish between the lumen of the gastrointestinal tract and the glandular lumen. However, this is to be expected because of a very similar appearance between these histological components. As well as this, we only used small image tiles for developing our algorithm and therefore contextual information to empower the segmentation is limited. In future work, we may incorporate a larger input size to provide additional context to the algorithm. A potential downside of our algorithm is that the model is quite large and therefore can't typically use large batch sizes. This can have negative implications on processing times and can lead to poor estimation of moving averages during batch normalisation.

With a small modification, the network is able to precisely segment the lumen of the gland. We also observed that the segmentation is very accurate within benign glands. This is positive because we presumed that there may have been confusion between lumenal areas and areas containing goblet cells. After performing this segmentation, lumenal features can be used to empower current automated classification methods, that are limited to features extracted from solely the gland object. We also observe that the additional segmentation of the lumen leads to an overall superior gland segmentation. This suggests that the lumen can provide additional cues to help increase the overall performance of gland instance segmentation.

In future work we will develop our proposed method for successful and fast whole-slide image processing. Therefore, we aim to adapt our method such that it can process a WSI in a short amount of time, whilst maintaining a similar level of accuracy. Our current method utilises RTS for uncertainty map generation. Although this uncertainty map is very informative, we must develop an approach that doesn't require ensembling if we plan to efficiently process the WSI in a short amount of time. As well as this, we will develop an effective pre-processing pipeline to ensure non-informative regions are not processed. On another note, it must be made clear that this algorithm is currently limited to colon cancer because of the data that it was trained on. The work could be extended such that we are able to segment the glands within other tissue, given that we have sufficient data.

In this chapter, we presented a minimal information loss dilated network for gland instance segmentation in colon histology images. The proposed network retains maximal information during feature extraction that is very important for successful gland instance segmentation. Furthermore, in order to segment glands of various sizes, we use *atrous* spatial pyramid pooling for effective multi-scale aggregation. To incorporate uncertainty within our framework, we apply random transformations to images during test time. Taking the average of this sample leads to a superior segmentation, whilst simultaneously allowing us to visualise areas of ambiguity. Furthermore, we propose an object-level uncertainty score that can be used for assessing whether to discard predictions with high uncertainty. We also highlight the generalisability of our method by processing whole-slide images from a different centre with high accuracy. As an extension, we show how our proposed method can be adapted such that it simultaneously segments the gland lumen and the gland object. We observe that our method obtains state-of-the-art performance in the MICCAI 2015 gland segmentation challenge and on a second independent colorectal adenocarcinoma dataset.

Chapter 5

Exploiting Rotational Symmetry in Histology Images

The recent advances in the analysis of Haematoxylin & Eosin (H&E) stained whole-slide images (WSIs) can largely be attributed to the rise of digital slide scanning [137]. In particular, Convolutional Neural Networks (CNNs) leverage the prior knowledge that images have translational symmetry and utilise a weight sharing strategy, which guarantees that a translation of the input will result in a proportional translation of the features. This property, known as *translation equivariance*, is an inherent property of the CNN and removes the need to learn features at all spatial locations, significantly reducing the number of learnable parameters. In certain image analysis applications, where there is no global orientation, it is desirable to extend this property of equivariance beyond translation to also rotation. One such example is the field of computational pathology (CPath) where important image features can appear at any orientation (Fig. 5.1). Therefore, we should be able to learn those features, regardless of their orientation. In the absence of rotation-equivariance, data augmentation is typically used, where multiple rotated copies of the WSI patches are usually introduced to the network during the training process. However, the augmentation strategy requires many more parameters in order to learn weights of different orientations. Instead, encoding rotational symmetry as a prior knowledge into current deep learning architectures by enforcing rotation-equivariance requires fewer parameters and leads to an overall superior discriminative ability. Also, rotation-equivariant CNNs typically converge quicker because the network does not need to spend time learning different filter orientations.

CPath is ripe ground for the utilisation of rotation-equivariant models, yet most models fail to incorporate this prior knowledge into the CNN architectures.

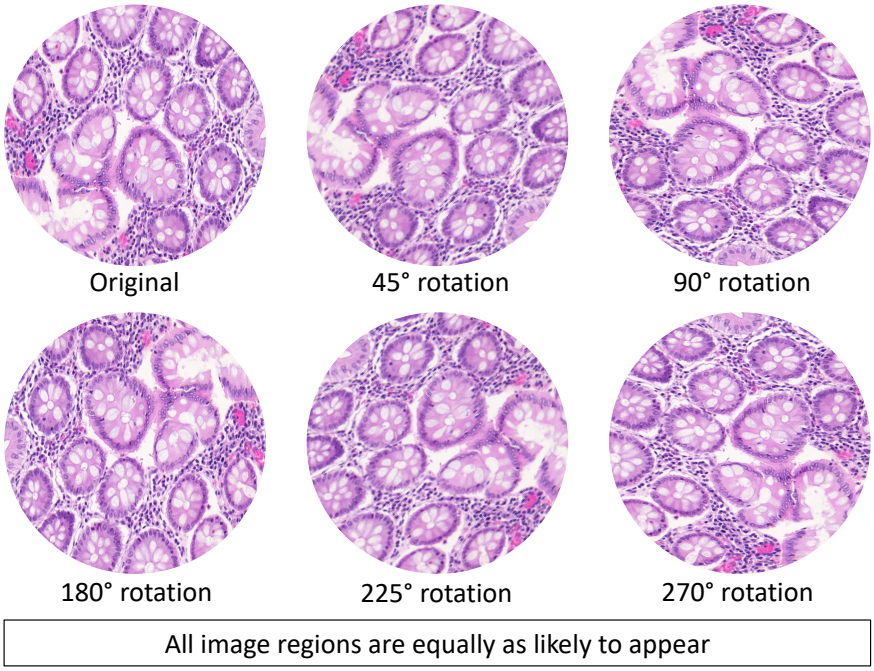


Figure 5.1: Cropped circular regions from a whole-slide image. Each orientation is equally as likely to appear.

Inspired by recent developments in the study of rotation-equivariant CNNs [32, 156, 110, 90], we propose two models in this chapter. First, we propose Rota-Net which is a dual-branch rotation-equivariant fully convolutional neural network (FCN) for simultaneous gland and lumen segmentation in colon histology images. Rota-Net uses the concept of group-convolutions which rotate the filters by multiples of 90° in addition to translation to enable the network to be equivariant to this group of symmetries. This approach can only perform exact rotation of standard filters if 90° rotations are used. Therefore, to overcome this challenge we then propose Dense Steerable Filter based CNNs (DSF-CNNs) that integrate steerable filters [47] with the group-convolution [32] and a densely connected framework [71] for superior performance. Each filter is defined as a linear combination of circular harmonic basis filters, enabling exact rotation and significantly reducing the number of parameters compared to standard filters. The main contributions of this chapter are listed as follows:

- A fully convolutional neural network for simultaneous gland and lumen segmentation that is equivariant to translations and 90° rotations.
- A Dense Steerable Filter CNN that is equivariant to translations and rotations

with a finer resolution than 90° by integrating steerable filter group convolutions within a densely connected network.

- The first thorough comparison of multiple rotation-equivariant for CPath.
- We demonstrate state-of-the art performance across multiple histology image datasets.

5.1 Related Work

5.1.1 CNNs for Translation Equivariance

Images can contain numerous symmetries and therefore patterns may appear at various spatial positions and orientations. Recent methods [79] have shown that these symmetries can be detected, yet in this work we focus on how symmetries can be leveraged as a *prior knowledge* to increase the performance of image recognition algorithms. Pioneered by LeCun *et al.* in 1994 [95], CNNs inherently incorporate translation symmetry in images and achieve translation equivariance by re-using filters at all spatial locations. Therefore, a shift of the input leads to a proportional shift of the filter responses. This design drastically reduces the number of required parameters because features do not need to be learned independently at each location. Since the increase in computing power and the development of algorithms that assist network optimisation [72] CNNs have become deeper [66, 70], leading to current state-of-the-art performance in numerous image recognition tasks [41, 99]. As a result of the success of deep learning, CNNs have since been widely used in CPath for various tasks including: gland segmentation [57, 27]; nucleus segmentation [61, 113, 88]; mitosis detection [9]; cancer type prediction [60] and cancer grading [14, 128]. Yet, unlike translation, CNNs do not behave well with respect to rotation because this symmetry is not built into the network architecture.

5.1.2 Exploiting Rotational Symmetry

Rotating the data: It is well known that histology images have no global orientation and therefore standard practice is to apply rotation augmentation to the training data [143]. This improves performance, but requires many parameters and is therefore prone to overfitting. Also, there is no guarantee that CNNs trained with rotation augmentation will learn an equivariant representation and generalise to data with small rotations [16]. To reduce the variance of predictions of multiple orientations, test-time augmentation (TTA) can be used [111]. However, with TTA

inference time scales linearly with the number of augmented copies. TI-Pooling [91] utilises multiple rotated copies of the input in a twin network architecture, where a pooling operation over orientations is performed to find the optimal canonical instance of the input images for training. However, like TTA, TI-Pooling is computationally expensive.

Rotating the filters: Cohen & Welling [32] pioneered group equivariant CNNs (G-CNNs), where the convolution was generalised to share weights over additional symmetry groups beyond translation. However, they limited the filter transformation to 90° rotations and horizontal/vertical flips to ensure exact transformations on the 2D pixel grid. Veeling *et al.* [146] showed that these G-CNNs can be used to improve the performance of metastasis detection in breast histology images. Furthermore, Linmans *et al.* [100] and Graham *et al.* [58] extended the application of the G-CNNs proposed by Cohen & Welling to pixel-based segmentation in histology images, highlighting an improved performance over conventional CNNs. The symmetries of a square grid are limited to integer translations extended by the dihedral group of order 8 (4 reflections and 4 rotations). To counter the limitation of working wih square grids in the G-CNN, Hoogeboom *et al.* [68] used hexagonal filters. However, this strategy requires images to be resampled on a hexagonal lattice, which is an additional overhead. Instead of using exact filter rotations, Bekkers *et al.* [21] and Lafarge *et al.* [90] applied G-CNNs to several medical imaging tasks by rotating filters with bilinear interpolation. Therefore, this method was not restricted to rotations by multiples of 90°, but may introduce interpolation artefacts. Oriented response networks [170] use active rotating filters during the convolution that explicitly encodes location and orientation information within the feature maps.

The aforementioned methods carry forward the feature maps for each orientation throughout the network. Instead, Marcos *et al.* [110] converted the output of multiple convolutions with rotated filter copies to a vector field by considering the magnitude and angle of the highest scoring orientation at every spatial location, leading to more compact models. To help overcome the issue of inexact filter rotation, the method only considered parameters at the centre of each filter and therefore required larger filters and consequently more parameters.

Rotating the feature maps: Dieleman *et al.* proposed a method similar to the G-CNN, but instead of rotating the filters, the feature maps were rotated. This design choice has no effect on the equivariance, yet any rotation that is not a multiple of 90° may suffer from interpolation artefacts.

Steerable filters: CNNs that encode rotation-equivariance are typically only equivariant to *discrete* rotations. Cohen & Welling [33] first proposed steerable

CNNs and described a general mathematical theory that applies to both continuous and discrete groups. To achieve full 360° equivariance, Worrall *et al.* [159] used the concept of steerable filters [47] and constrained the weights to be complex circular harmonics. Cheng *et al.* [30] propose a rotation-equivariant CNN, named RotDCF, that decomposes filters over joint steerable bases across the space and the group geometry simultaneously. Weiler *et al.* [156] learned steerable filters as a linear combination of atomic basis filters, which enabled exact filter rotation within G-CNNs. Then, these steerable filters were used within the group convolution to enable the network to be equivariant to rotation. Weiler & Cesa [155] then performed an extensive comparison of rotation equivariant models using steerable filters.

5.2 Mathematical Framework

In this chapter we present the key mathematical concepts used in our framework. We first describe images, filters and feature maps as functions. We introduce steerable filters and describe the group-convolution (G-convolution) operation, which is performed with either standard or steerable filters. This operation leads to G-equivariance. Below, we deal with a single filter at a time, although the method actually needs a whole filter bank to be used. We follow the method described by Weiler *et al.* [156], but we use a slightly different formulation.

5.2.1 Images and feature maps as functions

We model an image as a map $f : \mathbb{C} \cong \mathbb{R}^2 \to \mathbb{R}$ with compact support[1]. Let $\mathcal{F}$ be the vector space over $\mathbb{R}$ of all $f : \mathbb{C} \to \mathbb{R}$, with compact support, and let $\mathcal{F}_{\mathbb{C}}$ be the vector space over $\mathbb{C}$ of all functions $f : \mathbb{C} \to \mathbb{C}$ with compact support.

We denote by $\mathsf{SE}(2)$ the group of isometries of the plane, omitting reflections. Each element of $\mathsf{SE}(2)$ can be written in the form $z \mapsto e^{i\theta}z + b$, where $z, b \in \mathbb{C}$ and $\theta \in \mathbb{R}$. If $g \in \mathsf{SE}(2)$ and $f \in \mathcal{F}$, we define $g.f \in \mathcal{F}$ by:

$$(g.f)(z) = f(g^{-1}(z)) \text{ for } z \in \mathbb{C}. \tag{5.1}$$

The same definition is used for $g.f : \mathbb{C} \to \mathbb{C}$ when $f \in \mathcal{F}_{\mathbb{C}}$.

[1]The *support* of f is the smallest closed subset of $\mathbb{C}$ containing $\{z \in \mathbb{C} \mid f(z) \neq 0\}$.

5.2.2 Steerable functions and filters:

The additive group of real numbers $\mathbb{R}$ acts on $\mathbb{C}$ by rotations keeping 0 fixed. By (5.1), it acts linearly on $\mathcal{F}$ (and on $\mathcal{F}_{\mathbb{C}}$):

$$f^{\theta}(z) = f(e^{-i\theta}z) \text{ for } f \in \mathcal{F},\ \theta \in \mathbb{R}.$$

We define $V(f) \subset \mathcal{F}_{\mathbb{C}}$ to be the complex vector subspace spanned by the orbit $\left\{f^{\theta} \mid \theta \in \mathbb{R}\right\}$. If $V(f)$ is a finite dimensional vector space, we say that f is *steerable*.

Theorem: A necessary and sufficient condition for $\psi \in \mathcal{F}_{\mathbb{C}}$ to be steerable is that there should exist an integer $A \geq 0$, and radial profile functions $R_k : [0, \infty) \to \mathbb{C}$ for $k \in \mathbb{Z}$ and $-A \leq k \leq A$, such that, in polar coordinates:

$$\psi(r, \varphi) = \sum_{k=-A}^{A} R_k(r) e^{ik\varphi}, \tag{5.2}$$

where some or all of the radial profile functions R_k may be identically zero. To ensure that ψ has compact support, each R_k is assumed to have compact support.

If ψ satisfies (5.2), then $V(\psi)$ is clearly finite dimensional. The reverse implication takes a bit longer to argue, but easily follows from standard theorems in Group Representation Theory[2].

Fig. 5.2 is a graphical representation of basis harmonic filters that appear in (5.2).

Real Version: In practice we will work with steerable real-valued filters. Since a real-valued steerable filter ψ is also a complex-valued steerable filter, we can apply (5.2) to obtain, in the same notation:

$$\psi(r, \varphi) = \mathrm{Re}\left(\sum_{k=-A}^{A} R_k(r) e^{ik\varphi}\right).$$

Now $\mathrm{Re}(z) = (z + \bar{z})/2$. It follows that we can write instead (but the radial profiles change):

$$\psi(r, \varphi) = \mathrm{Re}\left(\sum_{k=0}^{A} R_k(r) e^{ik\varphi}\right) \tag{5.3}$$

where $R_0 : [0, \infty) \to \mathbb{R}$ and, for $k > 0$, $R_k : [0, \infty) \to \mathbb{C}$.

[2]For full mathematical rigour, the theorem requires the additional hypothesis that, for each r, ψ is a continuous function of φ. See also [139] for more technical details.

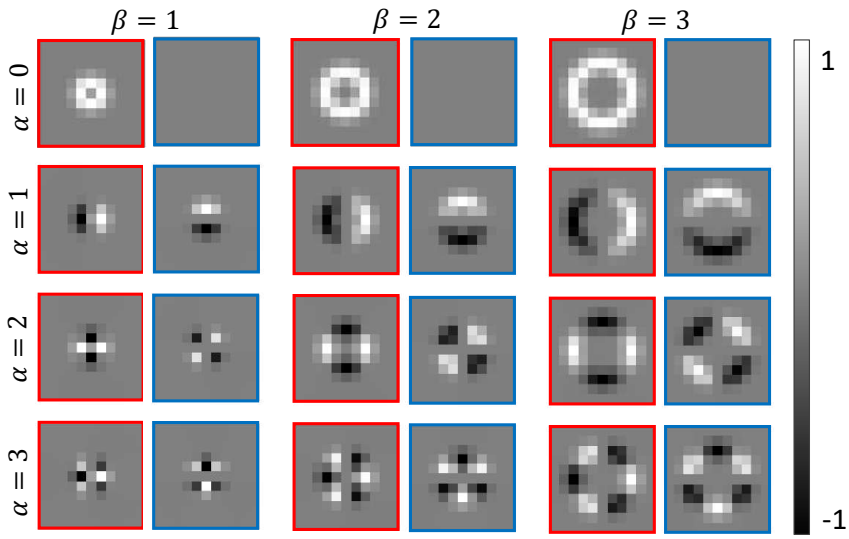


Figure 5.2: Example circular harmonic basis filters sampled on the 11×11 square grid. Red and blue borders denote the real and imaginary parts respectively. Each pair of images comes from a single term $R_k(r)e^{ik\theta}$ in (5.2). In this Fig., the particular radial profile functions R_k are all Gaussians, as they are in our proposed model. These Gaussians have mean/mode/max at j. The integer k specifies the frequency.

5.2.3 Feature maps modelled on a group:

Following the pioneering work of Cohen and Welling [32] and of Weiler *et al.* [156], we explain the changes to the architecture of CNNs, required to express rotation-equivariance.

We fix an integer $n > 0$. We use the symbol $\rho_{u,\theta}$ to denote the euclidean transformation given by

$$\rho_{u,\theta}(z) = e^{i\theta}z + u, \tag{5.4}$$

where $u \in \mathbb{C}$ and $\theta = 2\pi s/n$, for some integer s with $0 \leq s < n$. Let $G \subset \mathsf{E}(2)$ be the subgroup of all such transformations.

Let U be a group, with two subgroups U_1 and U_2. U is said to be a *semidirect product* of U_1 with U_2, denoted by $U_1 \rtimes U_2$, if there are projections $p_1 : U \to U_1$ and $U \to U_2$—this means that $p_1|U_1$ and $p_2|U_2$ are both identity maps—such that p_2 is a homomorphism with kernel U_1, and $p_1 \times p_2 : U \to U_1 \times U_2$ is an bijection, but, in general, not an isomorphism of groups. The importance of this concept in the study of equivariant CNNs was first pointed out in [32], and there is a systematic study [155].

G has two important subgroups, namely

$$C_n = \{\rho_{0,\theta} \mid \theta = 2\pi s/n, 0 \leq s < n\}, \tag{5.5}$$

a cyclic subgroup of order n consisting of all rotations in G keeping $0 \in \mathbb{C}$ fixed and

$$T = \{\rho_{u,0} \mid u \in \mathbb{C}\} \cong \mathbb{C},$$

consisting of all translations of $\mathbb{C}$. We define the group

$$C'_n = \{\theta \mid \theta = 2\pi s/n, 0 \leq s < n\}, \tag{5.6}$$

with group law addition mod 2π. Clearly, $C_n \cong C'_n$. We also use $\{e\} \cong C_1$ to denote the trivial group with one element.

The bijection

$$\mathit{\Pi} : G \to \mathbb{C} \times C'_n \text{ defined by } \mathit{\Pi}(\rho_{u,\theta}) = (u, \theta) \tag{5.7}$$

gives G the semidirect product structure $G = T \rtimes C_n$. We impose on $\mathbb{C} \times C'_n$ a product metric that is the same as the usual Euclidean metric on $\mathbb{C}$, and is any convenient fixed metric on the finite discrete space C'_n. The bijection $\mathit{\Pi}$ is then used to impose a metric on G, so that $\mathit{\Pi}$ becomes an isometry. $\mathit{\Pi}$ does not preserve the group structure, unless $n = 1$..

As a metric space G is the disjoint union of the n right cosets

$$\mathbb{C}_\theta = T\rho_{0,\theta} = \{\rho_{u,\theta} \mid u \in \mathbb{C}\} \subset G \text{ for } \theta \in C'_n, \tag{5.8}$$

such that each coset is isometric to $\mathbb{C}$.

A *G-feature map* is defined to be a function $f : G \to \mathbb{R}$, with compact support.

5.2.4 $\mathcal{G}$-convolutions:

We generalize the concept of a convolution to a G-convolution, that maps one G-feature map to another.

We give the definition of $\mathcal{G}$-convolution, where $\mathcal{G}$[3] is a group with a measure $\mu_{\mathcal{G}}$—this means that, given $f : \mathcal{G} \to \mathbb{R}$, we can form the integral denoted by $\int_{g\in\mathcal{G}} f(g)\, d\mu_{\mathcal{G}}$ or $\int_{g\in\mathcal{G}} f(g)\, dg$. We will stick to the *unimodular* case, which is gen-

[3]We use $\mathcal{G}$ instead of G because we have reserved the name G for the particular group defined in Subsection 5.2.3 and $\mathcal{G}$ denotes an arbitrary group.

eral enough for all cases of interest in this paper. The word *unimodular* means that we can change the dummy variable g in the integral to g^{-1}, or gh or hg ($h \in \mathcal{G}$ constant), without changing the value of the integral.

Given maps $f : \mathcal{G} \to \mathbb{R}$ and $\psi : \mathcal{G} \to \mathbb{R}$, we define their *$\mathcal{G}$-convolution* $(f *_{\mathcal{G}} \psi) : \mathcal{G} \to \mathbb{R}$ by

$$\begin{aligned}(f *_{\mathcal{G}} \psi)(g) &= \int_{h \in \mathcal{G}} f(gh^{-1})\psi(h)\, dh \\ &= \int_{h \in \mathcal{G}} f(h)\psi(h^{-1}g)\, dh \text{ for } g \in \mathcal{G}.\end{aligned} \tag{5.9}$$

The first equality is a definition, whereas the second follows by a change of variable.

$\mathcal{G}$-convolution is automatically $\mathcal{G}$-equivariant. To see this, note that, for any $\alpha \in \mathcal{G}$,

$$\begin{aligned}(\rho_\alpha(f) *_{\mathcal{G}} \psi)(g) &= \int f(\alpha^{-1}gh^{-1})\psi(h)\, dh \\ &= (f *_{\mathcal{G}} \psi)(\alpha^{-1}g) = (\rho_\alpha(f *_{\mathcal{G}} \psi))(g).\end{aligned}$$

It follows that

$$\rho_\alpha(f) *_{\mathcal{G}} \psi = \rho_\alpha(f *_{\mathcal{G}} \psi). \tag{5.10}$$

5.2.5 Hidden layer G-convolutions and G-filters

By a *G-filter*, we mean a function $G \to \mathbb{R}$. Formally this is the same as a G-feature map. However, in an implementation of these ideas, a G-feature map will turn out to be a discrete object, specified by a collection of matrices, whereas a G-filter retains its identity as a function. This is what enables exact rotation of a G-filter by an arbitrary angle.

In order to define G-convolutions, we need a measure on the space G, as described for $\mathcal{G}$ in Subsection 5.2.4. The measure μ_G on G is given by using the usual euclidean (area) measure on each $\mathbb{C}_\theta \cong \mathbb{C}$. Note that (G, μ_G) is *unimodular* (term defined in Subsection 5.2.4) because rotation is measure preserving on the plane. Integration of a function $f : G \to \mathbb{R}$, with respect to μ_G, is carried out by first integrating each of the n functions $f|\mathbb{C}_\theta \cong \mathbb{C} \to \mathbb{R}$ and adding the n resulting terms.

We now define an "*atomic steerable planar filter*", which is not learned, but defined and does not change during training (see (5.13)). Instead our network learns the complex coefficients used in a complex linear combination of the atomic steerable planar filters.

For each non-negative integer j, we define $\tau_j : [0, \infty) \to \mathbb{R}$ to be a Gaussian, with mode at j, as

$$\tau_j(r) = \exp(-|r - j|^2/2\sigma^2) \text{ for } j \geq 0,\ r \geq 0. \tag{5.11}$$

Let j and k be non-negative integers. By a *atomic steerable planar filter*, we mean a map $\psi_{jk} : \mathbb{C} \to \mathbb{C}$ defined by

$$\psi_{jk}(u) = \tau_j(|u|)e^{ik \arg(u)}. \tag{5.12}$$

If, in addition, $\lambda \in C'_n$, we define the *atomic steerable G-filter* $\psi_{jk\lambda} : G \to \mathbb{R}$ by

$$\psi_{jk\lambda}(\rho_{u,\theta}) = \begin{cases} 0 & \text{if } \lambda \neq \theta \\ \tau_j(|u|)e^{ik(\arg(u)-\theta)} & \text{if } \lambda = \theta. \end{cases} \tag{5.13}$$

From (5.12)

$$\psi_{jk\lambda}(\rho_{u,\theta}) = e^{-ik\theta}\psi_{jk}(u) \text{ if } \theta = \lambda, \tag{5.14}$$

which is ψ_{jk} rotated by angle θ.

Any finite complex linear combination of atomic steerable G-filters, $\sum_{j,k,\lambda} w_{jk\lambda}\psi_{jk\lambda}$, is again a steerable G-filter. In our framework, we plan to convolve each G-feature map with the real part of such a sum. By (5.9) the result of such a convolution is another G-feature map. The complex numbers $w_{jk\lambda}$ are weights in the network, determined by the network during training and each $w_{jk\lambda}$ gives rise to two real weights. We will initially restrict to a single term in the finite sum, in order to keep the formulas uncluttered, and then add them together.

Let $f : G \to \mathbb{R}$ be a G-feature map. From (5.9), we have the formula

$$\begin{aligned} &(f *_G \operatorname{Re}(w_{jk\lambda}\psi_{jk\lambda}))\,(\rho_{z,\theta}) = \\ &\qquad \int_{\rho_{u,\varphi} \in G} f(\rho_{u,\varphi}) \cdot \operatorname{Re}(w_{jk\lambda}\psi_{jk\lambda}(\rho_{v,\beta}))\, d\mu_G, \end{aligned} \tag{5.15}$$

where $\rho_{v,\beta} = \rho_{u,\varphi}^{-1}\rho_{z,\theta}$, so that $v = e^{-i\varphi}(z - u)$ and $\beta = \theta - \varphi$. From (5.12) and (5.13),

$$\psi_{jk\lambda}(\rho_{v,\beta}) = \begin{cases} 0 & \text{if } \lambda \neq \beta = \theta - \varphi \\ e^{-ik\varphi} \cdot \psi_{jk}(z - u) & \text{if } \lambda = \beta = \theta - \varphi. \end{cases} \tag{5.16}$$

Writing $f_\varphi(u) = f(\rho_{u,\varphi})$, we obtain from (5.15) and (5.16)

$$\begin{aligned}
&\left(f *_G \mathrm{Re}(w_{jk\lambda}\psi_{jk\lambda})\right)(\rho_{z,\theta}) \\
&= \quad \mathrm{Re}\left(w_{jk\lambda} \cdot e^{-ik(\theta-\lambda)} \cdot (f_{\theta-\lambda} * \psi_{jk})\right)(z) \\
&= \quad \left(f_{\theta-\lambda} * \mathrm{Re}(w_{jk\lambda} \cdot e^{-ik(\theta-\lambda)}\psi_{jk})\right)(z).
\end{aligned} \tag{5.17}$$

If we add over $\lambda \in C'_n$, then we can substitute $\varphi = \theta - \lambda$ and add over $\varphi \in C'_n$, since θ is fixed in (5.17). Adding over j, k and φ, we obtain

$$\begin{aligned}
&\left(f *_G \mathrm{Re}(\sum_{jk\lambda} w_{jk\lambda}\psi_{jk\lambda})\right)(\rho_{z,\theta}) \\
&= \quad \sum_{jk\varphi} \left(f_\varphi * \mathrm{Re}\left(w_{jk(\theta-\varphi)} \cdot e^{-ik\varphi}\psi_{jk}\right)\right)(z)
\end{aligned} \tag{5.18}$$

which recovers the same result as (10) in [156]. We have ignored the fact that there are normally many channels (G-feature maps) in the domain and many channels in the range. Each pair (channel in domain, channel in base) needs its own G-filter, so each such pair gives rise to different weights.

5.2.6 The input layer G-convolution

The input to network is an image that can be thought of as a map $f : \mathbb{C} \to \mathbb{R}$, which we compose with $P : G \to \mathbb{C}$ given by $P(\rho_{u,\theta}) = u$, to obtain $f \circ P : G \to \mathbb{R}$. By (5.17), we have

$$\begin{aligned}
&\left((f \circ P) *_G \mathrm{Re}(w\psi_{jk\lambda})\right)(\rho_{z,\theta}) = \\
&\quad \mathrm{Re}((w_{jk\lambda} \cdot e^{ik\lambda}) \cdot e^{-ik\theta} \cdot (f * \psi_{jk})(z)
\end{aligned}$$

Since $w_{jk\lambda}$ is a complex scalar that the network has to estimate, λ adds no new information and we dispense with it. We then sum over all terms, obtaining a simplified version of (5.18).

$$\begin{aligned}
&\left((f \circ P) *_G \mathrm{Re}\left(\sum_{jk} w_{jk}\psi_{jk}\right)\right)(\rho_{z,\theta}) \\
&= \quad \left(f * \mathrm{Re}\left(\sum_{jk} w_{jk} \cdot e^{-ik\theta} \cdot \psi_{jk}\right)\right)(z).
\end{aligned} \tag{5.19}$$

This gives a principled derivation of Equation (8) in [156]. In particular, our proof of G-equivariance (see (5.10)) works equally well for input layer and hidden layer G-convolutions. See Fig. 5.4 for a graphical illustration of the method.

5.2.7 Sampling and the discrete case

The above formulas assume that the functions involved are continuous. But a computer is a finite machine, so we need to work with discrete data, and this involves sampling.

Sampling planar steerable filters: In the computer, a planar feature map is represented by a matrix, not by a continuous function. According to (5.18) and (5.19), we need to convolve this matrix with the real part of a complex linear combination of atomic planar filters, ψ_{jk}. Now ψ_{jk} is a function, not a matrix—this is exactly what allows rotation of the filter through an arbitrary angle. On the other hand, convolution with a matrix requires a matrix, not a function. We therefore have to sample the atomic filters ψ_{jk}, and their rotations through angles $2\pi s/n$ for $0 \leq s < n$, at the integer points $a + ib$, where a and b are integers. We then perform a weighted linear combination of the sampled filters and apply (5.18) or (5.19). As the Nyquist Sampling Theorem suggests, for a fixed size of steerable filter, aliasing may occur unless one bounds the frequencies used from above. In line with Weiler & Cesa [155], we use frequencies up to $k = 0, 2, 3, 2$ for $j = 0, 1, 2, 3$ in all 7×7 steerable basis filters. Using larger filters enables higher frequencies before aliasing, yet leads to an increase in computation time and may lead to overfitting.

Sampling G-filters: As in the case of planar convolution just discussed, our formulas need to be reinterpreted when the various component pieces of a hidden layer G-convolution are formulated as arrays of dimension 3 or higher, rather than as functions. For example a G-feature map has been defined as a function $G \to \mathbb{R}$, and we need to explain how a function on the continuous group G is represented in the computer by n matrices.

As shown in (5.8), G as a metric space is the disjoint union $\bigcup_{\theta\in C'_n} \mathbb{C}_\theta$ of n copies of $\mathbb{C}$, with its usual euclidean metric. For each $\theta \in C'_n$ (see (5.8)) we define

$$\mathbb{Z}_\theta = \{\rho_{a+ib,\theta} \mid a, b \in \mathbb{Z}\} \subset \mathbb{C}_\theta. \tag{5.20}$$

Each point of $\mathbb{C}_\theta$ is within a distance $1/\sqrt{2}$ of some point in the lattice $\mathbb{Z}_\theta$. It is therefore reasonable to use, as a G-feature map,

$$f : \bigcup_{\theta\in C'_n} \mathbb{Z}_\theta \to \mathbb{R}. \tag{5.21}$$

Analogously to the notation just before (5.17), we write $f_\theta = f|\mathbb{Z}_\theta$. The domain is infinite, but since f is assumed to have compact support, we need only record the values of f at a finite number of elements of G. In this way, a G-feature map is replaced by n real matrices all of the same size.

We have also defined a G-filter as a function $G \to \mathbb{R}$. This is also sampled on $\bigcup_{\theta \in C'_n} \mathbb{Z}_\theta$. When learning the complex coefficients $w_{jk\lambda}$ that appear in (5.15), the values of j and k are limited for the reasons just explained for the planar situation, namely to avoid aliasing and overfitting.

5.3 Methods

In this section we present two methods: Rota-Net and Dense Steerable Filter (DSF) CNNs, which both incorporate rotational symmetries into their architecture. Rota-Net is developed as an initial experiment to assess whether the incorporation of rotational symmetry into the convolution leads to an improved performance in CPath. For Rota-Net, we applied it to the specific task of gland and lumen segmentation as a proof of concept. Then, after analysis of Rota-Net, we developed DSF-CNNs that enabled rotation with a finer resolution by learning steerable filters. We applied our DSF-CNN to the tasks of gland segmentation, nuclear segmentation and breast tumour classification. Below, we provide a description of each of the models.

5.3.1 Rota-Net

Network Architecture

The overall network architecture, as shown in Figure 5.3, is based on the fully convolutional network [103] architecture, with residual blocks [66] for efficient gradient propagation. The network first downsamples features with max-pooling by a factor of 16, which increases the size of the receptive field, before upsampling with bilinear interpolation to increase the spatial resolution. The main components of Rota-Net can be summarised as: input G-convolution layer, G-residual blocks, upsampling and a G-mean-pooling layer. Below we provide a description of each of the components of the network.

Input layer G-convolution: Throughout Rota-Net, we utilise G-convolutions with standard filters that are translated across the input and rotated by 90°. In the first layer, each filter is a conventional 3×3 filter that is translated over the input. However, the convolution process is repeated for each orientation of the filter to give 4 orientation dependent outputs. Therefore, the input is a function on the plane

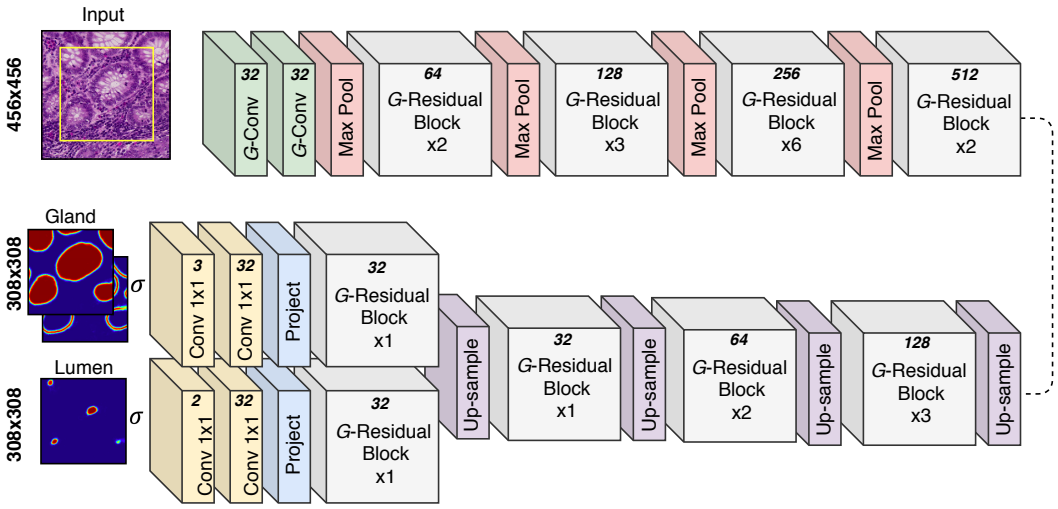


Figure 5.3: Rota-Net architecture. The yellow box within the input denotes the part of the image considered at the output. The number at the top of each operation denotes the number of feature maps produced per filter orientation. Note, for group operations, this is the number per orientation of the kernel (4 orientations in the $p4$ group). σ is the softmax operation.

$\mathbb{Z}^2$, but the output feature maps are a function on the group G. We then observe that if the input is rotated, the feature maps also rotate but undergo an additional channel permutation. Therefore, this is the action of the group G.

Input G-residual-blocks: Because the sum of two rotation-equivariant feature maps is also rotation-equivariant, residual blocks are well suited within this network design. Within our framework, a G-residual-block consists of multiple G-residual-units, where each unit consists of two 3×3 G-convolutions and a shortcut connection. All G-convolutions within each G-residual-block has an input and an output both on the group G. Therefore, filters are also on the group G and thus perform a channel permutation with rotation to mimic the group action. All G-convolutions within the network are followed by rotation-equivariant batch normalisation, where moments are aggregated per group, and a ReLU.

Upsampling: After downsampling the features, we use bilinear interpolation to upsample feature maps. Each time, we upsample by a factor of 2 followed by a G-residual-block. We use valid convolution in the upsampling branch which leads to the output being smaller than the input, thus reducing boundary artefacts when processing neighbouring image patches. Similar to U-Net [121], we utilise skip connections with addition to incorporate low level features at the output of the network. In the same vein as the residual unit, this addition is rotation equivariant. The network splits after the final upsampling operation, where each branch is

subsequently devoted to either gland or lumen segmentation.

G-mean-pooling: Because feature maps within the network are functions on the group G, features need to be projected back to a function on the plane at the output of the network. We achieve this by defining the *projection layer* that takes the average over the 4 orientations. This operation is followed by two consecutive planar 1x1 convolution operations to obtain the final output.

5.3.2 Dense Steerable Filter CNN

Network Architecture

The main building blocks of our proposed rotation-equivariant DSF-CNN[4] are: input layer G-convolution layer; steerable filter G-dense-blocks, upsampling and a G-max-pooling layer. Below, we build on the theoretical explanation in Section 5.2 to describe the separate components of our proposed approach.

Input Layer G-convolution: Up to the G-pooling operation, all convolutions within our network are steerable G-convolutions, as described in Section 5.2.5. Therefore, we pre-define a set of circular harmonic basis filters using (5.2) and sample the filters on the square grid, as can be seen in Fig. 5.2. Then, we learn how to linearly combine these atomic basis filters to generate steerable filters and consider only the real part for our convolution filter, as shown in (5.3). The input layer steerable G-convolution maps an image $f : \mathbb{C} \to \mathbb{R}$ to some G-feature map $h : G \to \mathbb{R}$. Each G-feature map is determined by its restriction h_θ to each coset $\mathbb{C}_\theta \cong \mathbb{C}$. Specifically, we create n rotated copies of each steerable filter and independently convolve the filters with the input to produce n feature maps (or a single G-feature map). Planar rotation of each filter is performed using (5.14) and can observed in Fig. 5.6. The input layer G-convolution is demonstrated in Fig. 5.4, where the convolution between the input and the steerable filter bordered in red produces the output also bordered in red. Now, when the input is rotated by an angle $\frac{2\pi s}{n}$, with integers $0 \leq s < n$, and the input layer G-convolution is performed, the feature maps undergo a planar rotation by angle $\frac{2\pi s}{n}$, but in addition shift s positions.

G-dense-blocks: To enable efficient gradient propagation, encourage feature re-use and to improve overall performance, we use dense connectivity [70] between G-convolutions in hidden layers of the network. Each hidden layer steerable G-convolution maps a G-feature map $f : G \to \mathbb{R}$ to some G-feature map $h : G \to \mathbb{R}$. We can explain this mapping in terms of the restrictions of f and h to cosets. Be-

[4]Model code: `https://github.com/simongraham/dsf-cnn`

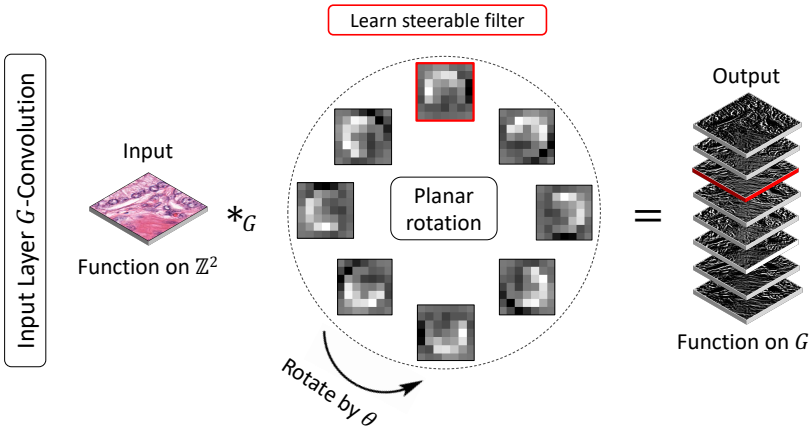


Figure 5.4: Illustration of the input layer G-convolution, mapping an image $f : \mathbb{C} \to \mathbb{R}$ to a G-feature map $h : G \to \mathbb{R}$. A single steerable planar filter, learned by the network, is rotated n times and each rotated filter is convolved with the planar input f. This gives n planar feature maps, which combine to give a single G-feature map h. The image f is convolved with the red bordered planar filter to give the red bordered planar feature map in the stack on the right.

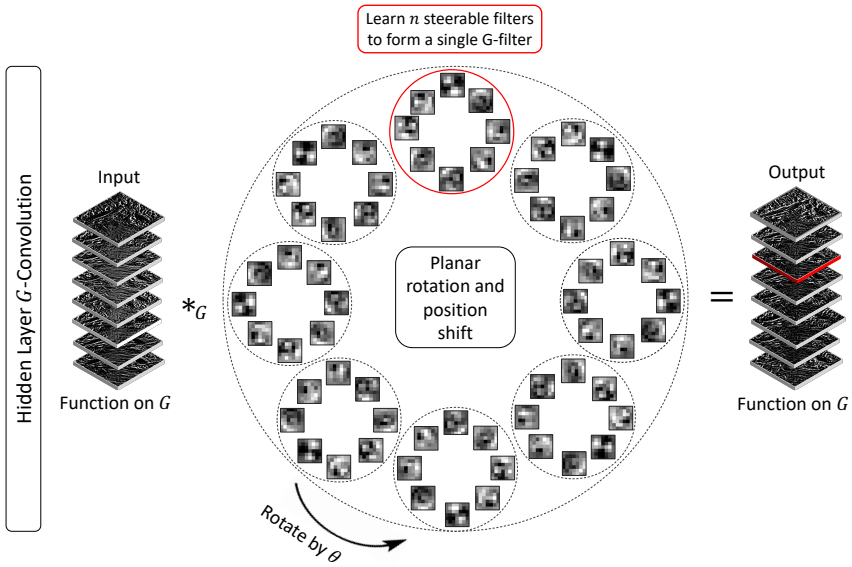


Figure 5.5: Illustration of the hidden layer G-convolution, mapping a G-feature map $f : G \to \mathbb{R}$ to a G-feature map $h : G \to \mathbb{R}$. The network learns a single steerable G-filter, which consists of n planar filters, displayed by placing them all in the same circle. Then, a single G-filter is rotated n times and each rotated G-filter is convolved with the input G-feature map f to generate a total of n planar feature maps or a single G-feature map. The convolution between the input f and the red circled G-filter gives the red bordered planar feature map on the right.

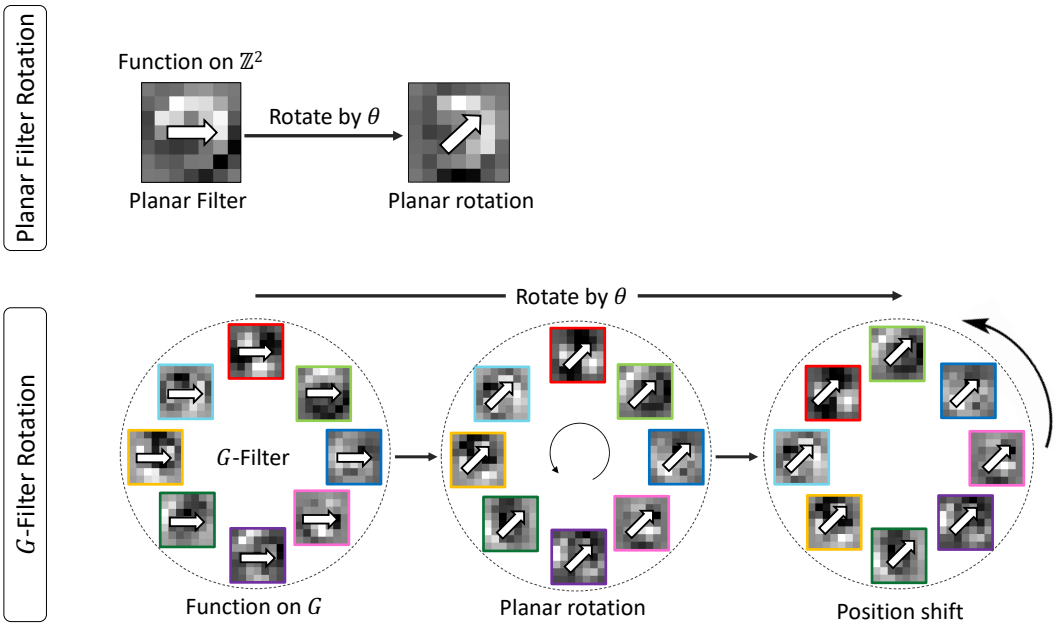


Figure 5.6: Planar filter and G-filter rotation. Planar filters are rotated in the conventional manner, whereas G-filters undergo and additional position shift after planar rotation. In the displayed example, both filters rotate by an angle $\theta = \frac{\pi}{4}$

cause the input to the hidden layer G-convolution is now a function on G, we must similarly ensure that our filters give a function on G. We rotate each G-filter to give n rotated copies and perform a convolution between the input G-feature map f and each filter orientation to produce n feature maps (or a single G-feature map h). When rotating these G-filters, an additional position shift must be performed, in line with the associated group action. In Fig. 5.5, $n = 8$ steerable planar filters are generated as shown by the red circle, forming a single G-filter. This G-filter is convolved with the input G-feature map to generate the output with the red border. We can see that each G-filter, consists of 8 planar filters that individually rotate and shift position as the entire G-filter is rotated. This rotation can be seen in Fig. 5.6, where the arrows show the orientation of each planar filter and the coloured borders are used to help visualise the position of each planar filter in the G-filter.

For each G-dense-block, the feature-maps of all preceding layers are concatenated to the input before performing the G-convolution. This increases the number of connections between layers, strengthening feature propagation. Specifically, each G-dense-block consists of k units. Each unit contains a 7×7 G-convolution followed by a 5×5 G-convolution that produce 14 and 6 orientation dependent feature maps respectively. After k units, the G-dense-block concludes by applying a final 5×5

G-convolution. All G-convolutions are followed by rotation-equivariant batch normalisation, where moments are aggregated per group rather than spatial feature map.

Upsampling: Similar to Rota-Net, we upsample feature maps with bilinear interpolation after feature extraction. For this, feature maps are upsampled sequentially by a factor of 2 and upsampling operations are followed by G-dense-blocks. Also, features from the encoder are added to the decoder with each upsampling operation to achieve a better performance [121].

G-max-pooling: At the output of the network, we transform each G-feature map f to a planar feature map, by taking the pointwise maximum of the n planar feature maps f_θ that constitute f. This operation ensures that the output of G-pooling is *invariant* to rotation of the input.

Classification: For our classification DSF-CNN, we initially perform the input layer steerable G-convolution followed by a hidden layer G-convolution. We then use 4 G-dense-blocks, where each block consists of 3,4,5 and 6 dense units. After every G-convolution layer we use a group-equivariant batch normalisation that aggregates moments per group rather than spatial feature map and ReLU non-linearity. Before every G-dense-block, we perform spatial max-pooling to decrease the dimensions of the feature maps. After the final G-dense-block, we perform G-pooling and then apply 3 1×1 classical convolution operations to get the final output.

Segmentation: We extend our DSF-CNN to the task of segmentation by up-sampling feature maps after the final G-dense-block in the aforementioned classification CNN. Specifically, we up-sample by a factor of 2 with bilinear interpolation and then utilise a G-dense-block. This is repeated until the spatial dimensions of the original image are regained. From the deepest layer of the up-sampling branch, each dense-block contain 4, 3 and 2 units. In line with U-Net [121], we also use skip connections to propagate information from the encoder to the decoder. After the feature maps have been up-sampled, we use a single hidden layer G-convolution, which is followed by G-pooling such that the resulting feature map is a function on $\mathbb{C}$. Finally we use 2 1×1 classical convolutions to obtain the output, where we segment both the object and the contour to help separate touching instances. For nuclear segmentation, we additionally predict the eroded nuclei masks which are used as markers in marker-controlled watershed.

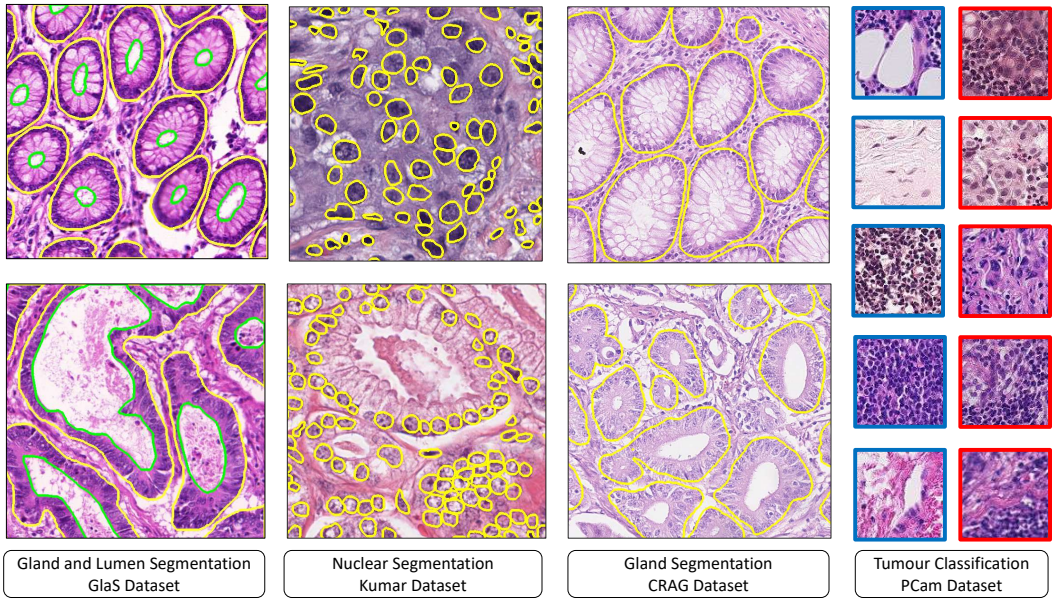


Figure 5.7: Image regions from the four datasets: GlaS [135], Kumar [88], CRAG [57] and PCam [146]. Yellow and green boundaries show the pathologist annotation of nuclei/glands and lumen respectively. Blue and red borders denote non-tumour and tumour image patches.

5.4 Experiments and Results

5.4.1 The Four Datasets

We use the following four publicly available histology image datasets:

Breast tumour classification: PCam [146] is a dataset of 327K image patches of size 96×96 pixels at 10× extracted from the Camelyon16 dataset [19], containing 400 H&E stained breast WSIs. Each image patch was labelled as tumour if the central region (32×32) contained at least one tumour pixel as given by the original annotation [19].

Multi-tissue nucleus segmentation: The Kumar dataset [88] contains 30 image tiles of size 1,000×1,000 from seven organs (6 breast, 6 liver, 6 kidney, 6 prostate, 2 bladder, 2 colon and 2 stomach) of The Cancer Genome Atlas (TCGA) database acquired at 40× magnification. Within each image, the boundary of each nucleus is fully annotated.

Colorectal gland segmentation: The CRAG dataset [57] consists of 213 H&E images mostly of size 1,512×1,516 pixels taken from 38 WSIs acquired at 20× of colorectal adenocarcinoma (CRA) patients. It is split into 173 training images and 40 test images with different cancer grades with pixel-based gland annotation.

Colorectal gland and lumen segmentation: The GlaS dataset [135] consists of 85 training (37 benign and 48 malignant) and 80 test images (37 benign and 43 malignant) regions extracted from 16 H&E stained WSIs at 20×. The test images are split into test sets A and B, where A was released one month before the challenge deadline and B was released on the final day of the challenge. Images are mostly of size 775×522 pixels and all training images have associated instance-level segmentation ground truth that precisely highlight the gland and lumen boundaries.

5.4.2 Evaluation Metrics

For tumour classification, we calculated the area under the receiver operating characteristic curve (AUC) to assess the binary classification performance. For gland/lumen segmentation, we employed the same quantitative measures that were used in the GlaS challenge [135]. These metrics consist of F_1, DICE and Hausdorff distance at the object level and assess the quality of instance segmentation. For nuclear segmentation, we report the binary DICE and panoptic quality (PQ). Here, the binary DICE assesses the ability of the method to distinguish nuclei from the background, whereas PQ provides insight into the quality of instance segmentation.

5.4.3 Experimental Overview

Recently, there has been a growing number of proposed CNNs that achieve rotation-equivariance [32, 156, 110, 21, 159], yet there is lack of comprehensive evaluation of the various methods for the analysis of histopathology images. We perform a thorough comparison of various rotation-equivariant CNNs and demonstrate the effectiveness of the proposed model. Specifically, we compare a baseline CNN with H-Nets [159], VF-CNNs [110], G-CNNs with standard filters [32, 21] and G-CNNs with steerable filters [156] and assess the impact of increasing the number of filter rotations in each model. After gaining an insight into the performance of the different rotation-equivariant models, we then quantify the performance of Rota-Net on the task of gland and lumen segmentation and DSF-CNN on the tasks of breast tumour classification, nuclear segmentation and gland segmentation. The rest of this section is split into three parts:

- Comparative analysis of rotation equivariant models
- Quantitative and visual evaluation of Rota-Net
- Quantitative and visual evaluation of DSF-CNN

5.4.4 Comparative Analysis of Rotation-Equivariant Models

Comparative Model Description

Baseline models: For the task of breast tumour classification, we implement a baseline CNN for comparison with the aforementioned rotation-equivariant models. The model consists of a series of convolution, batch normalisation, non-linear and spatial pooling operations, which are then followed by three 1×1 convolutions to obtain the final output, denoting the probability of an input patch being tumour.

For the tasks of gland and nuclear segmentation we leverage the fully convolutional neural network architecture, which allows us to use the same model architecture, irrespective of the input size. The encoder of the baseline segmentation model uses the same architecture as the baseline classification CNN. Then a series of up-sampling and convolution operations are used to regain the spatial dimensions of the original image. In line with U-Net, we use skip connections to incorporate features from the encoder, but utilise summation as opposed to concatenation. At the output of the network we perform segmentation of the object and the contour and additionally predict the eroded masks for nuclear segmentation.

Rotation-equivariant models: To assess the performance of various rotation-equivariant approaches, we modify the baseline models, but keep the fundamental architecture the same. The main difference between different models is how the filters are rotated, how many filter orientations are considered and how the convolution operation is performed.

Aside from H-Nets, each rotation-equivariant model considers 4, 8 and 12 filter orientations. H-Nets encode full 360° equivariance within the model and therefore filters do not need to be explicitly rotated. When applying rotation to a filter with an angle that is a multiple of $\frac{\pi}{2}$, the rotation is *exact* because the output can still be represented on the square grid. However, any other rotation may give interpolation artefacts and therefore may have negative implications for rotation-equivariance. Therefore, in line with Marcos *et al.* [110] and Lafarge *et al.* [90], for both the VF-CNN and standard G-CNN, we apply circular masking to the filters when using the groups C_8 and C_{12}. However, this masking still leads to inevitable interpolation artefacts in the centre of the filter. Steerable filters as defined by (5.2) do not suffer from interpolation artefacts and, therefore, circular masking is not needed.

In all comparative experiments for rotation-equivariance, we fix each filter to be of size 7×7. We used a larger filter than typically used in modern CNNs because this size ensures that we can construct a good basis set for steerable filter

generation, with reasonable frequency content and reduced aliasing.

For fair comparison, we ensure that the number of parameters is similar between different models. For both standard and steerable G-CNNs, the number of parameters increases with the size of the group, if we fix the number of filters in each layer. This is because one feature map is produced per orientation of the filter, which increases the number of required filters in the subsequent layer. To maintain the same number of parameters as the baseline CNN, we divide the number of filters in each layer of the standard G-CNN by $\sqrt{n}$, where n is the number of orientations in the group. Steerable G-CNNs learn k parameters (or $k/2$ complex parameters) for each filter, where typically $k < K^2$. Therefore, the number of filters in each layer of a steerable G-CNN should be divided by $\frac{k\sqrt{n}}{K^2}$. Instead of carrying forward all orientations throughout the network, VF-CNNs collapse the orientation dependent feature maps to two feature maps, representing magnitude and angle. Therefore, the VF-CNN requires more filters in the next layer, but the number of parameters stays constant irrespective of the size of the group. To ensure the same number of parameters as the baseline CNN, for all group sizes we divide the number of filters in each layer of VF-CNNs by $\frac{4}{3}$. Each H-Net filter is constrained to be a complex circular harmonic, parameterised by N radial terms and a single phase offset term. Also, the number of parameters is dependent on the maximum frequency m of the filters. Specifically, in H-Nets frequencies in the range $[-m, m]$ are considered, equating to a total of $M = 2m + 1$ frequency terms. Therefore, to ensure a similar number of parameters as the standard CNN, we multiply the number of filters in each layer of a H-Net by $\frac{K^2}{M \cdot (N+1)}$.

In all models, we down-sample with max-pooling, but for VF-CNNs and H-Nets we use a modified pooling strategy, based on the magnitude of the feature maps. Similarly, when using both VF-CNNs amd H-Nets, we do not incorporate the angle information when using batch normalisation (BN) and non-linear activation functions; otherwise the angles may change important information about relative and global orientations. For G-CNNs, we use a modified BN that aggregates moments per group rather than spatial feature map.

To verify our implementations of the various rotation-equivariant networks, we cross-checked the performance of each model against reported benchmarks on the rotated MNIST dataset [92] before applying them to the histology datasets. These results are summarised in Table B.2.

Table 5.1: Tumour classification results on the PCam dataset [146]. All models have a similar parameter budget. The superscript associated with H-Net denotes the maximum frequency used.

Method	**Group**	**Parameters**	**AUC**
CNN	$\{e\}$	$564K$	0.947
H-Net[1] [159]	$\mathsf{SO}(2)$	$553K$	0.934
H-Net[2] [159]	$\mathsf{SO}(2)$	$542K$	0.939
VF-CNN [110]	C_4	$556K$	0.949
VF-CNN [110]	C_8	$556K$	0.951
VF-CNN [110]	C_{12}	$556K$	0.953
G-CNN [32]	C_4	$561K$	0.964
G-CNN [21, 90]	C_8	$557K$	0.968
G-CNN [21, 90]	C_{12}	$557K$	0.962
Steerable G-CNN [156]	$\{e\}$	$553K$	0.963
Steerable G-CNN [156]	C_4	$546K$	0.969
Steerable G-CNN [156]	C_8	$565K$	0.971
Steerable G-CNN [156]	C_{12}	$545K$	0.969

Quantitative Results of Rotation-Equivariant Models

Tumour classification: We report comparative results of different rotation-equivariant models on the PCam dataset at the top of Table 5.1. We observe that H-Nets do not perform as well as the baseline CNN for the task of tumour classification. Despite this, we observe that we are able to increase the performance when incorporating higher frequency filters in the network, but the performance is still not comparable to conventional CNNs. This may suggest that constraining the filters in this way may not be optimal for detecting complex features in histology. VF-CNNs marginally outperform the conventional CNN, where we observe that increasing the number of filter rotations leads to a slight improvement in performance. When we utilise the group convolution, with filter rotation as performed by Bekkers *et al.* [21] and Lafarge *et al.* [90], we see an improved performance when using up to 8 filter orientations. This gain in performance can be attributed to incorporating our prior knowledge of rotational symmetry into the network. To ensure that we maintain a similar number of parameters, we need to reduce the number of feature maps at each layer when the size of the group is increased. This may explain the drop in performance when using 12 filter orientations. When using steerable filters, but with no filter rotation, we observe an improved performance over conventional CNNs, highlighting the benefit of learning a linear combination of basis filters, rather than standard filters. Then, as we increase the size of the group to 4 and 8 orientations

Table 5.2: Gland segmentation results on the CRAG [57] dataset, where all models have a similar parameter budget.

Method	**Group**	**Params**	**Obj F_1**	**Obj Dice**	**Obj Haus ↓**
CNN	$\{e\}$	$984K$	0.793	0.809	246.0
G-CNN [32]	C_4	$982K$	0.833	0.856	170.4
G-CNN [21, 90]	C_8	$988K$	0.837	0.866	157.4
G-CNN [21, 90]	C_{12}	$979K$	0.818	0.834	192.2
Steerable G-CNN [156]	$\{e\}$	$981K$	0.811	0.848	175.9
Steerable G-CNN [156]	C_4	$984K$	0.837	0.869	164.8
Steerable G-CNN [156]	C_8	$989K$	0.861	0.888	139.5
Steerable G-CNN [156]	C_{12}	$976K$	0.855	0.870	156.2

Table 5.3: Nuclear segmentation results on the Kumar [88], where all models have a similar parameter budget.

Method	**Group**	**Params**	**B-Dice**	**PQ**
CNN	$\{e\}$	$984K$	0.767	0.447
G-CNN [32]	C_4	$982K$	0.793	0.490
G-CNN [21, 90]	C_8	$988K$	0.811	0.519
G-CNN [21, 90]	C_{12}	$979K$	0.814	0.534
Steerable G-CNN [156]	$\{e\}$	$981K$	0.791	0.510
Steerable G-CNN [156]	C_4	$984K$	0.809	0.542
Steerable G-CNN [156]	C_8	$989K$	0.818	0.543
Steerable G-CNN [156]	C_{12}	$976K$	0.820	0.558

we see an improvement in the performance. We also observe that using steerable filters rather than standard filters within the G-convolution gives a better result.

Gland segmentation: We compare the performance of the different rotation-equivariant models for gland segmentation on the CRAG dataset in the top part of Table 5.2. For this experiment, when comparing different rotation-equivariant approaches, we choose to only assess the performance of conventional CNNs, standard G-CNNs and steerable G-CNNs. This is because our previous experiment on breast tumour classification indicates that G-CNNs are capable of achieving a superior result over competing rotation-equivariant approaches. Similar to our observations for breast tumour classification, we see that increasing the group size within the group convolution leads to an increase in performance, but the best performance is achieved when using 8 filter orientations. For this task, using steerable filters in the group convolution led to the best performance.

Nuclear segmentation: We report the comparative results of different

rotation-equivariance methods for nuclear segmentation on the Kumar dataset in the top part of Table 5.3. Similar to above, we compare conventional CNNs with both standard and steerable G-CNNs. Here, we see that all rotation-equivariant approaches show a significant improvement over standard CNNs and we see an improvement when increasing the number of filter orientations to 12 in all models. Once again, we observe that the steerable G-CNNs for segmentation of nuclei are superior to standard G-CNNs that use bilinear interpolation during filter rotation.

We evaluate the performance of our proposed method with several state-of-the-art approaches in the bottom part of Table 5.3. In particular, HoVer-Net [61], CIA-Net [169], Micro-Net [119] and DIST [113] have been purpose-built for the task of nuclear segmentation and, therefore, provide a competitive benchmark. The proposed DSF-CNN once again achieves the best performance compared to other methods for both binary DICE and panoptic quality, on par with the state-of-the-art HoVer-Net method, while requiring a fraction of the parameter count.

5.4.5 Visualisation of Features and Output

In Figs. 5.8 and 5.8 we visualise the features and the corresponding outputs as we rotate the input with angle increments of $\frac{\pi}{4}$ (8 in total) for both the baseline CNN and C_8-steerable G-CNN. Specifically, we analyse the properties of both CNNs trained for the tasks of gland and nuclear segmentation. To observe the feature map transformation with rotation of the input, we analyse two sets of feature maps in both CNNs: *Feature Map A* at the output of the 2nd convolution and *Feature Map B* at the output of the convolution after the final up-sampling operation. Similarly, we observe how the output probability map transforms when the input is rotated.

To analyse this, we feed each image orientation into the network to obtain a set of feature maps and output probability maps. Then, after rotating features and probability maps back to their original orientation, we compute the pixel-wise variance map of the features and the output to see how they change with rotation of the input. G-CNN feature maps are a function on G and therefore we visualise a single planar feature map within the group. For the rotation-equivariant model, we observe that there is a near-negligible variance between the features of each input orientation. On the other hand, there is much higher variance between the features of standard CNNs after input rotation. This implies that the rotation-equivariant CNN successfully learns an equivariant feature representation. Also, there is a lower variance between the predictions of multiple input orientations for the rotation-equivariant CNN as compared to the standard CNN. Thus, the rotation-equivariant CNN behaves as expected with rotation of the input, which is a particularly desirable

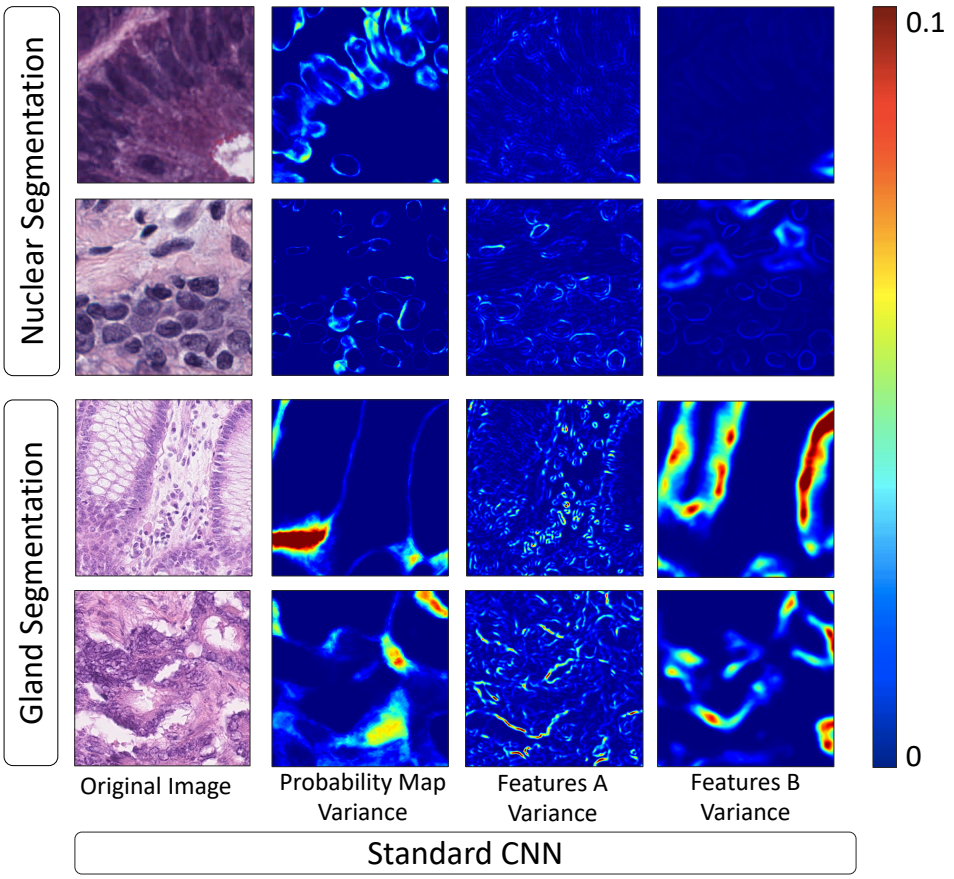


Figure 5.8: Variance between the predictions and features of a standard CNN for multiple orientations of the input. The original image is rotated with steps of $\frac{\pi}{4}$ to give 8 orientations and each copy is passed through the network to enable variance calculation. Features A and B are located at the beginning and end of the network respectively.

property when training CNNs with histology image data. It must be noted that features learned by conventional CNNs are highly complex and it is very difficult to infer the relationship between learned features and input rotation. Nonetheless, we demonstrate that rotation-equivariant CNNs have a predictable transformation with input rotation, making them more stable than conventional CNNs.

5.4.6 Evaluation of Rota-Net

Quantitative Results of Rota-Net

To quantify the performance of Rota-Net, we first perform an ablation study to assess the contribution of the G-convolution that incorporates rotation-equivariance.

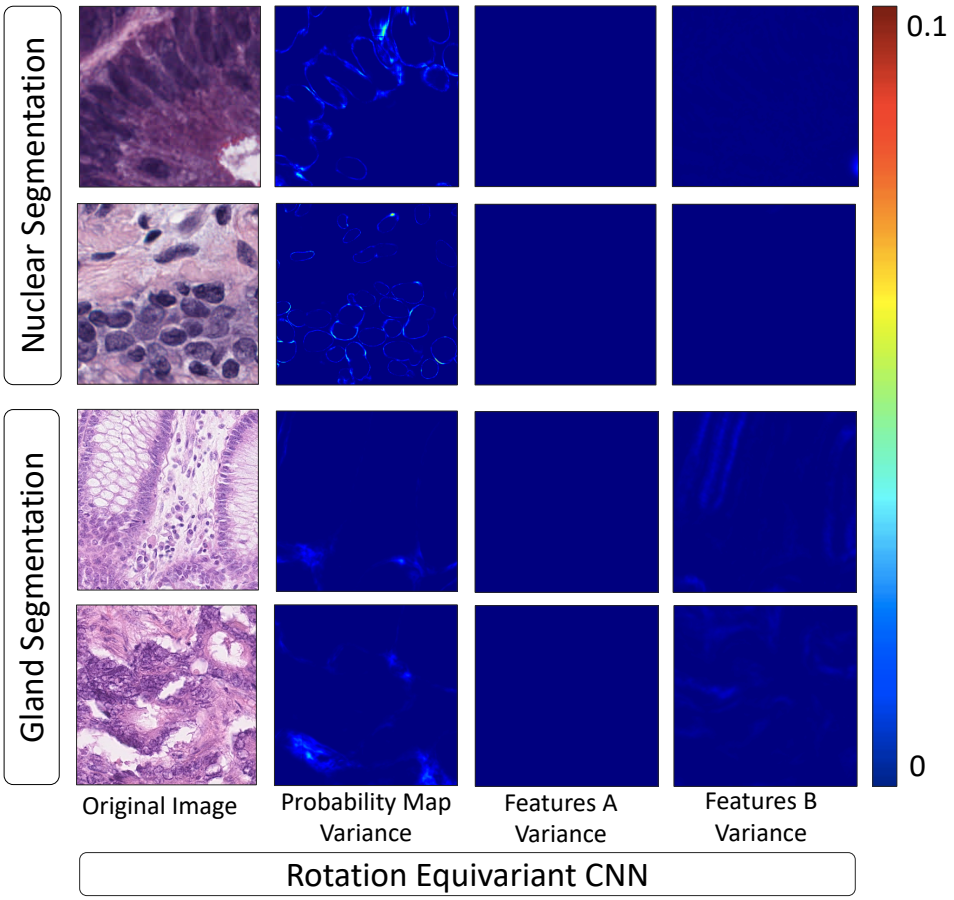


Figure 5.9: Variance between the predictions and features of a rotation-equivariant CNN for multiple orientations of the input. The original image is rotated with steps of $\frac{\pi}{4}$ to give 8 orientations and each copy is passed through the network to enable variance calculation. Features A and B are located at the beginning and end of the network respectively.

Then, we compare the performance of our proposed method to recent top-performing models. Table 5.4 highlights the contributions of the various network components in Rota-Net. It is evident that using the rotation-equivariant approach with group convolutions improved the performance. This therefore reflects the findings from Section 5.4.4 that rotation-equivariant methods lead to better performance in histology image analysis. This performance is further improved when the contours are considered for effective gland separation. Compared to the baseline network, we reduce the number of kernels in each layer of the rotation-equivariant network by a factor of two to maintain a similar number of parameters. This is in line with the explanation given in Section 5.4.4 .

Table 5.5 shows comparative results for simultaneous gland and lumen seg-

Table 5.4: Ablation study. RE denotes rotation equivariant network. RE$^+$ denotes rotation equivariant network, utilising a multi-class strategy at the gland output.

	$\mathbf{F_1}$ **Score**		**Obj. Dice**		**Obj. Hausdorff**		
Method	**Gland**	**Lumen**	**Gland**	**Lumen**	**Gland**	**Lumen**	**Params**
Baseline	0.905	0.715	0.899	0.739	50.29	73.36	70.6M
RE	0.916	0.789	0.913	0.807	46.00	57.49	71.3M
RE$^+$	**0.920**	**0.831**	**0.919**	**0.824**	**40.99**	**49.17**	71.3M

Table 5.5: Comparative results for simultaneous gland and lumen segmentation. All networks are converted to a dual-branch architecture, where the network splits after the final upsampling operation. Note, for conciseness we only evaluate on test set A.

	$\mathbf{F_1}$ **Score**		**Object Dice**		**Object Hausdorff**	
Method	**Gland**	**Lumen**	**Gland**	**Lumen**	**Gland**	**Lumen**
Rota-Net	**0.920**	**0.831**	**0.919**	**0.824**	**40.99**	**49.17**
U-Net [121]	0.857	0.643	0.846	0.725	86.63	70.59
FCN-8 [103]	0.800	0.735	0.820	0.762	99.98	68.80

Table 5.6: Comparative results for gland segmentation using Rota-Net.

	$\mathbf{F_1}$ **Score**		**Object Dice**		**Object Hausdorff**	
Method	**Test A**	**Test B**	**Test A**	**Test B**	**Test A**	**Test B**
Rota-Net	**0.920**	0.824	**0.919**	**0.849**	**40.99**	**95.72**
MILD-Net [57]	0.914	**0.844**	0.913	0.836	41.54	105.89
Multichannel B [161]	0.893	0.843	0.908	0.833	44.13	116.82
Micro-Net [119]	0.913	0.724	0.906	0.785	49.15	133.98
CUMedVision2 [26]	0.912	0.716	0.897	0.781	45.418	160.347
Freidburg2 [121]	0.870	0.695	0.876	0.786	57.09	148.47

mentation. For effective evaluation, we compare with a modified U-Net [121] and FCN-8 [103] where, in a similar fashion to Rota-Net, the branches split after the final upsampling operation. We observe that our proposed approach performs significantly better than both competing approaches and is able to simultaneously segment both glands and lumen with high accuracy.

In Table 5.6 we compare the gland segmentation performance of our proposed approach with recent top performing methods. I particular, the current state-of-the-art approach is MILD-Net that was presented in Chapter 4. We observe that our proposed Rota-Net achieves the best performance in five out of six metric and therefore exceeds the previous top performing approach for gland segmentation on the GlaS dataset.

Visual Results of Rota-Net

Figure 5.10 displays some visual results of the proposed method compared to the ground truth. We also display some areas of interest, shown by the black boxes in Figure 5.10(b) and (c), where the algorithm successfully segments lumen, but is missed by the pathologist. It is important to note that the proposed approach makes one prediction per pixel and no patch overlap is used during processing, whereas other approaches may make multiple predictions per pixel. For example, MILD-Net merges overlapping predictions and also uses a test-time augmentation strategy.

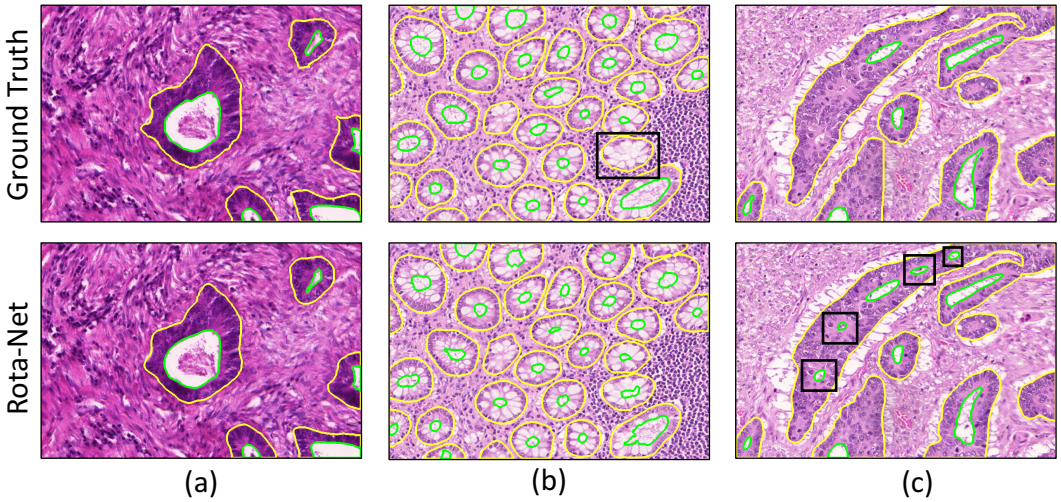


Figure 5.10: Visual results of gland and lumen segmentation using Rota-Net. Yellow and green boundaries denote gland and lumen boundaries respectively. Black boxes show areas of interest.

5.4.7 Evaluation of DSF-CNN

Quantitative Results of DSF-CNN

Tumour classification: In Table 5.7 we compare the performance of our proposed DSF-CNN with the $p4m$-DenseNet [146], which is the top performing method that was proposed with the introduction of the PCam dataset. This approach integrates the use of G-convolutions on, as proposed by Cohen & Welling [32], into a densely connected CNN [71]. Here, the network uses filter rotations by multiples of 90° and also uses reflections. This is denoted by D_4, which is the dihedral group containing 4 rotation and 4 reflection symmetries. In addition, we compare results to the commonly used ResNet-34 [66], ResNet-50 [66], DenseNet-121 [71] and DenseNet-169 [71]. Despite the small amount of parameters, we observe that our method achieves

the best performance with an AUC of 0.975, which is a promising improvement over the previous state-of-the-art.

Table 5.7: Comparison of DSF-CNN with state-of-the-art on the PCam dataset [146].

Method	**Group**	**Parameters**	**AUC**
ResNet-34 [66]	$\{e\}$	$21.3M$	0.942
ResNet-50 [66]	$\{e\}$	$23.5M$	0.948
DenseNet-121 [71]	$\{e\}$	$7.8M$	0.921
DenseNet-169 [71]	$\{e\}$	$13.3M$	0.920
$p4m$-DenseNet* [146]	D_4	$119K$	0.963
DSF-CNN **(Ours)**	C_8	$2.2M$	**0.975**

Table 5.8: Comparison of DSF-CNN with state-of-the-art on the CRAG dataset [57].

Method	**Group**	**Params**	**Obj F_1**	**Obj Dice**	**Obj Haus** $\downarrow$
FCN8 [121]	$\{e\}$	$134.3M$	0.796	0.835	199.5
U-Net [121]	$\{e\}$	$37.0M$	0.827	0.844	196.9
MILD-Net [57]	$\{e\}$	$83.3M$	0.869	0.883	146.2
Rota-Net [58]	$\{e\}$	$71.3M$	0.869	0.887	144.2
DSF-CNN **(Ours)**	C_8	$3.7M$	**0.874**	**0.891**	**139.5**

Table 5.9: Comparison of DSF-CNN with state-of-the-art on the Kumar dataset [88].

Method	**Group**	**Params**	**B-Dice**	**PQ**
FCN8 [103]	$\{e\}$	$134.3M$	0.797	0.312
SegNet [17]	$\{e\}$	$29.4M$	0.811	0.407
U-Net [121]	$\{e\}$	$37.0M$	0.758	0.478
Mask-RCNN [65]	$\{e\}$	$40.1K$	0.760	0.509
DIST [113]	$\{e\}$	$9.2M$	0.789	0.443
Micro-Net [119]	$\{e\}$	$192.6M$	0.797	0.519
CIA-Net [169]	$\{e\}$	$22.0M$	0.818	0.577
HoVer-Net [61]	$\{e\}$	$54.7M$	**0.826**	0.597
DSF-CNN **(Ours)**	C_8	$3.7M$	**0.826**	**0.600**

Gland segmentation: In Table 5.8, we compare our proposed approach with MILD-Net [57] and Rota-Net [58], which are top-performing gland segmentation methods and therefore can be appropriately used for performance benchmarking. As mentioned in the above section, like the $p4m$-DesneNet, Rota-Net makes

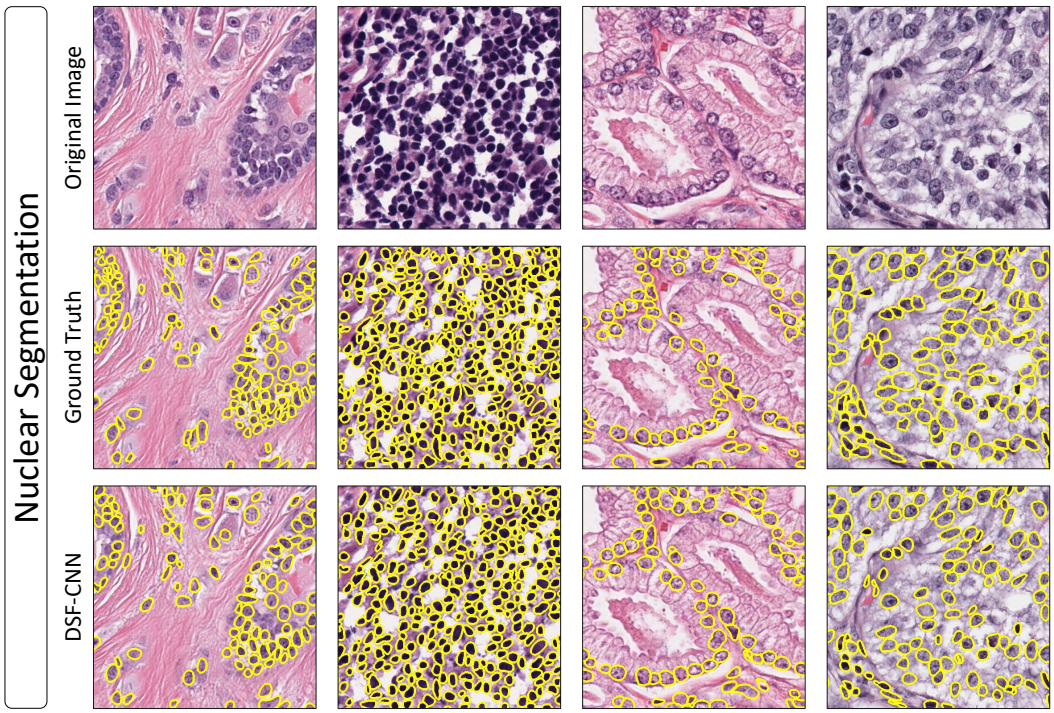


Figure 5.11: Visual results of nuclear segmentation using DSF-CNN. Yellow boundaries show either the pathologist annotation or predicted nuclei.

use of the standard G-convolution, but is limited to only 90° filter rotations. In addition, we compare with FCN8 and U-Net as they are two widely used CNNs for segmentation. We observe that our DSF-CNN achieves the best performance with a fraction of the parameter budget. Notably, our model has around 20 times fewer parameters than Rota-Net and MILD-Net.

Nuclear segmentation: We evaluate the performance of our proposed method with several state-of-the-art approaches in Table 5.9. In particular, HoVer-Net [61], CIA-Net [169], Micro-Net [119] and DIST [113] have been purpose-built for the task of nuclear segmentation and, therefore, provide a competitive benchmark. The proposed DSF-CNN once again achieves the best performance compared to other methods for both binary DICE and panoptic quality, on par with the state-of-the-art HoVer-Net method, while requiring a fraction of the parameter count.

Visual Results of DSF-CNN

In Figures 5.11 and 5.12 we show some visual results for nuclei and gland segmentation, where the yellow boundaries show either the pathologist annotation or the nuclei/gland predictions. We see that our algorithm is able to perform a good qual-

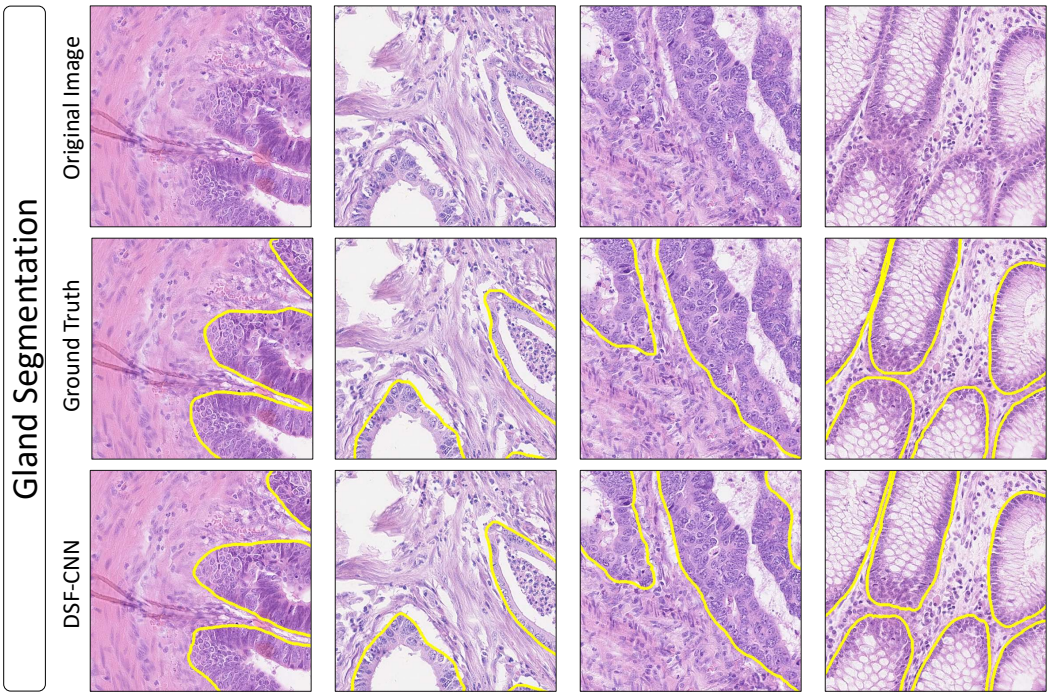


Figure 5.12: Visual results of gland segmentation using DSF-CNN. Yellow boundaries show either the pathologist annotation or predicted glands.

ity segmentation of the nuclei and glands, where the prediction closely resembles the GT. In particular, we see that our algorithm is able to successfully differentiate between touching nuclei and glands and hence can be appropriately used for subsequent object-based feature extraction in downstream analysis.

5.4.8 Implementation and Training Details

We implemented our framework with the open source software library TensorFlow version 1.12.0 [7] on a workstation equipped with two NVIDIA GeForce 1080 Ti GPUs. During training, data augmentation including flip, rotation, Gaussian blur and median blur was applied. For breast tumour classification, we fed the original patches of size 96×96 into the network. For gland, lumen and nuclear segmentation, we used patches of size 448×448 and 256×256 respectively. For tumour classification, we trained our model using a batch size of 32 and then used a batch size of 8 for segmentation models. We used cross-entropy loss for all tumour classification and gland/lumen segmentation models, whereas we used a combination of weighted cross-entropy and dice loss for nuclear segmentation. For all models, we trained using Adam optimisation with an initial learning rate of 10^{-3}, that was reduced

as training progressed. The network was trained with an RGB input, normalised between 0 and 1.

5.5 Discussion and Conclusions

Conventional CNNs do not behave as expected with rotation of the input, which is a particularly undesirable property in the field of computational pathology, where important features in histology images can appear at any orientation. Instead, rotation-equivariant CNNs build this prior knowledge of rotational symmetry within the network, such that features rotate in accordance with the input without explicitly learning features at various orientations. In this chapter, we propose two networks: Rota-Net and DSF-CNN. Rota-Net introduces the concept of rotation-equivariance and motivates its use within computational pathology by displaying improved results over conventional CNNs. Then, we enhance Rota-Net by proposing a densely connected steerable filter CNN that achieves state-of-the-art performance on the tasks of tumour classification, gland segmentation and nuclear segmentation with a fraction of the parameter budget of recent top-performing models. We conducted a thorough comparative analysis of various rotation-equivariant CNNs on the 3 tasks mentioned above. We showed that steerable filter group convolutions gave the best quantitative results on all three tasks, where 8 filter orientations consistently gave a strong performance. We visualised features within a rotation-equivariant model to demonstrate that they rotate with the input and therefore have a higher degree of feature map interpretability. Finally, we showed that rotation-equivariant models give more stable predictions with input rotation than regular CNNs do. In future work, we will consider incorporating additional symmetries into the group convolution, such as mirror and scale symmetries. This will further increase the interpretability of feature maps and may lead to an improvement in performance. Also, the exploration of further symmetries in histology images may help direct future research in computational pathology

Chapter 6

Conclusions and Future Directions

In this thesis, we presented a range of computational tools to facilitate the automatic analysis of cancerous tissue in H&E WSIs. We first addressed the challenge of dealing with large-scale WSIs in CPath and developed a pipeline for automatic NSCLC WSI classification. Then, we developed an algorithm for simultaneous nuclear segmentation and classification, followed by an algorithm for gland segmentation. Finally, we investigated rotation-equivariant CNNs for CPath and developed several models applied to the tasks of simultaneous gland and lumen segmentation, gland segmentation, nuclear segmentation and tumour classification. All of the machine learning methods described are supervised learning approaches, fundamentally based on convolutional neural networks.

Apart from our WSI classification pipeline, the majority of this thesis focused on the localisation of nuclei or other components, such as glands, within the tissue. It must be noted that localisation is typically not the end goal in CPath and further work is needed to integrate these algorithms into a structured pipeline. For example, the simultaneous segmentation and classification of nuclei enables subsequent downstream analysis of the nuclei within a WSI, opening up possibilities of further analysis of large-scale nuclear morphometry. Features can be directly extracted from segmented nuclei and used in an ML model to predict clinical outcome. First localising areas of interest and then utilising a set of known features is often referred to as a *bottom-up* approach and can provide greater *explainability* of WSI-level predictions. Of course, this approach is not limited to nuclei but can be applied to any localised structure within the tissue. For instance, gland segmentation can similarly be used as a prerequisite step before morphological feature extraction and

patient outcome prediction from the WSI. In particular, extracted features that reflect glandular aberrance [15] can provide an objective and explainable measure that can help overcome the challenge of subjectivity in visual assessment of gland formation.

Below we provide recommendations for how some of the work presented in this thesis may be extended and we discuss potential future directions.

6.1 Opportunities for Future Research

6.1.1 Simultaneous Segmentation and Classification of Nuclei

In Chapter 3 we proposed a method for simultaneous segmentation and classification of nuclei. Because we remove padding during the convolution in the upsampling branch (also known as valid convolution), the output is smaller than the size of the input. The size of the output also determines the maximum stride that can be used when processing patches in WSIs. Therefore, the smaller the output size of the network, the smaller the maximum stride will be, which consequently has an adverse effect on the total time to process each WSI. In our case, there is a significant difference in the input and output size (270×270 vs 80×80) and therefore WSI processing time will suffer. Segmenting nuclei within WSIs is typically done as an initial step before downstream analysis. Therefore, it is important to optimise this step to prevent unreasonably long processing times for the overall CPath pipeline. In future work we may increase the efficiency of our nuclear segmentation and classification algorithm to make it suitable for WSI processing. The first obvious adjustment would be to increase the size of the output to enable larger strides. Further work can also be spent on increasing the efficiency of the network. For instance, we may prune the filters of the CNN [96] identified as having a small effect on the output accuracy to effectively reduce the number of convolution operations in the model. The concept of *knowledge distillation* via teacher-student networks can also be used to decrease the size of the model. Here, the teacher would be our proposed network and the student would be a more compact version with fewer parameters. Then, the knowledge distillation scheme encourages the student network to make predictions that closely resemble the predictions made by the teacher network.

As mentioned in Section 3.5, horizontal and vertical maps are better suited to convex objects, which is why they perform particularly well for the task of nuclear segmentation. In future work, we may extend our concept of horizontal and vertical distance maps so that they can also be used for segmenting non-convex shaped structures.

As mentioned above, performing the segmentation is not the final step and an additional step is required to make a WSI-level prediction. One powerful group of methods that can leverage segmented nuclei to make slide-level predictions are Graph Convolutional Networks (GCNs). GCNs have the ability to integrate both morphological features and graph-level features that reflect the spatial relationship between instances. Not only does this area hold great promise in providing an interpretable and powerful predictor, but may additionally help overcome the challenge of fitting the entire WSI into the memory of the GPU to train DL models. In future work, we plan to use the output of our nuclear segmentation and classification network as input to a GCN to predict cancer diagnosis.

6.1.2 Gland Segmentation

In Chapter 4 we introduced a method for accurate gland instance segmentation. A major component of this approach is the use of dilated convolution that introduces sparsity in the kernel and thus increases the size of the receptive field during convolution. One area of further exploration would be the use of deformable convolution [40], where the rate of dilation is learned as opposed to being explicitly set. Within our model for gland segmentation, we use the concept of MIL units to help counter the loss of information caused by max-pooling. In future work it is important to perform a full quantitative analysis to thoroughly quantify the effect that information loss from downsampling has on the performance of CNNs. Uncertainty quantification can be helpful in highlighting where a model has difficulties in making a prediction. There will inevitably be areas in the tissue that are naturally hard to diagnose and we should not force our model to make a prediction in these areas. However, to obtain an output of uncertainty, our method requires multiple copies of the image to be fed through the network. It would be beneficial to develop an approach that instead can output an uncertainty map, whilst only using one input image.

In future work, we aim to leverage the performance of our gland segmentation approach and study the relationship between glandular morphology and patient outcome. It has already been shown that the level of glandular aberrance, as measured by the best-alignment metric [15], can be used to directly predict the grade of colorectal cancer. However, this analysis was performed on small image regions extracted from the WSI. Instead, we aim to perform a large-scale WSI-level investigation of glandular morphology and explore its relationship with a range of clinical parameters, such as grade, recurrence and survival.

6.1.3 Exploiting Symmetries in CNNs

Chapter 5 exploited the inherent rotational symmetry present in histology images by making conventional CNNs rotation-equivariant. However, there are additional symmetries that histology images possess that have not been explored in this thesis. For example, histology images also have reflection symmetry. In other words, flipping a histological image does not change its image content and will appear with the same probability as the original image. Therefore, a natural development would be to additionally incorporate reflection symmetries into the G-convolution; therefore making it equivariant to the Euclidean group $\mathsf{E}(2)$. Similar to rotations by multiples of 90°, flipping the kernel along its horizontal or vertical axis is exact and therefore we do not need to worry about interpolation artefacts. This is the case for both standard and steerable filters. When using flips in addition to rotation, $2n$ feature maps will be produced per filter. Therefore, if using a fixed parameter budget, it is important to balance the trade-off between the number of symmetries utilised in the network and the number of independent filters in each layer.

Another symmetry group that would be interesting to incorporate into our framework is scale symmetry. This is because components within the tissue may appear at different scales due to differences in pixel resolution between scanner manufacturers. A possible direction to achieve scale-equivariance would be to use the concept of dilated convolution, as has been done by Worrall and Welling [158].

In recent work [146], it has been shown that rotation-equivariant CNNs are more sample efficient than standard counterparts. This is because a rotation-equivariant method will be able to recognise features regardless of their orientation. This is an important characteristic, especially in the medical domain, where labelled data is hard to obtain. It would be interesting to further explore this claim and demonstrate the performance of our proposed DSF-CNN with different proportions of the input data.

6.1.4 Immunohistochemistry Analysis

In this thesis, we focused primarily on the analysis of H&E stained histology images. However, a pathologist is also required to analyse immunohistochemistry (IHC) slides that enable visualisation of antigen expression in the tissue. To quantify the level of expression, pathologists often assign a score to the slide that is usually done by visual assessment. This is naturally subjective, due to the difficulty in counting a huge amount of positively stained cells, and therefore computational algorithms show great potential in enabling a more accurate and reproducible quantification.

In future work, our proposed HoVer-Net can be used to segment positively stained nuclei in IHC WSIs, enabling accurate counting, morphological assessment and determination of expression levels of individual cells. This automatic analysis can also enable further analysis of regions with high expressions levels and can help clinicians associate the level of expression with tissue morphology.

Obtaining ground truth for the development of algorithms in computational pathology is difficult because it needs to be validated by domain experts. Also this task can even be challenging for experts, due to the difficulty in determining 3D structures in a single 2D cross-sectional view. IHC data can be leveraged to accurately classify individual cells, which can otherwise be difficult in H&E stained slides. Most available datasets are typically curated by examination of H&E tissue and consequently may be prone to inter-observer variation. Therefore, the field of computational pathology would largely benefit from the development of a large nuclear instance segmentation dataset, where the cells are categorised by IHC. Following this approach would also enable the categorisation of millions of cells within a slide. This strategy has been used for mitotic cell recognition [142], but extension to all cells within a series of WSIs would be advantageous for cell-based approaches in CPath.

6.1.5 Open Problems in Computational Pathology

Throughout this thesis, we described a selection of methods that aim to tackle some key tasks in computational pathology. Of course, there are numerous other applications not described in this thesis where CPath can be advantageous. For example, an exciting application within this domain is the prediction of genetic alterations from tissue stained with H&E [77, 78, 37], where usually additional genetic or immunohistochemical tests are needed. Therefore, computational tools developed for mutation prediction can help potentially reduce turnaround time and cost. However, especially for the prediction of certain genetic mutations, performance is still quite low and therefore more work needs to be done before we can consider integrating it within diagnostic pipelines.

Similar to this, there are various other applications that can benefit from the the rich feature representations that CPath algorithms are capable of extracting. For example, computational tools may not only be used to help improve the objectivity, reproducibility and accuracy of diagnosis of tissue samples, but may also help with the advancement towards precision oncology. Tissue samples contain an abundance of complex information that can be leveraged to more optimally predict appropriate patient treatment. It is an open problem on how to best use compu-

tational pathology to help inform oncologists' decisions, yet is an area that in the near future may receive a lot of attention, due to the positive impact it may have on patient outcome. However, as we move towards developing algorithms that can have a dramatic impact on patients' lives, we must also strive for model explainability that can help inform the final decision made by the oncologist.

All of the algorithms developed in this thesis relied on accurately labelled datasets, where in the case of segmentation tasks, pixel-level annotation was needed. With the rise of digital pathology, there is a growing amount of available data, yet it is not feasible to expect pixel-level annotation for all available slides. Instead, for certain tasks it can be preferable to leverage a single slide-level label that is usually readily available. For instance, this slide label can correspond to the grade or type of cancer. Weakly supervised approaches for WSI classification, that utilise only the slide label as ground truth, have shown recent success in computational pathology [23, 105, 106]. An open challenge within CPath is the development of weakly supervised approaches that may help to explain why certain predictions have been made by the accurate localisation of discriminative regions. This is particularly interesting for tasks, where the association between tissue morphology and the label is not completely clear. For example, such approaches may help pathologists understand which morphological features are responsible for certain genetic alterations. These morphological features can be explored by utilisation of some of the segmentation algorithms mentioned in this thesis.

6.2 Closing Remarks

In this thesis we proposed a range of algorithms that automatically analyse cancerous tissue in H&E histology images. Many of the algorithms that we developed focus on the accurate localisation of nuclei and glands within the tissue and therefore may serve as a strong prerequisite before subsequent downstream analysis in CPath. We also provided motivation for the use of rotation-equivariant CNNs for histology image analysis, where rotational symmetry exists on a global scale.

To leverage the strong performance of the models presented in this thesis, full integration into diagnostic CPath pipelines is necessary before deployment into clinical practice. However, before this can be done, various other important factors need to be considered. First of all, it is imperative to conduct a large-scale validation of the developed algorithms across multicentric cohorts to assess generalisability to unseen data. This data should also include WSIs acquired with scanners from a variety of different manufacturers and tissue prepared with varying protocol. Also,

it is inevitable that slides will often contain artefacts that may lead to incorrect diagnoses. For example, if dirt exists on the glass slide before scanning, then it may result in the entire WSI being out-of-focus. Another example of artefacts present in WSIs are pen markings that pathologists sometimes draw to circle regions of interest. Therefore, due to the above regions, it is essential to integrate a pre-processing step for quality-control before application of CPath algorithms. Another vital criterion for successful deployment of CPath algorithms is the development of an easy to navigate user interface that pathologists can seamlessly integrate into their routine workflow.

In the future, we believe that CPath will be a fundamental component of the digital pathology workflow and will be prove pivotal in the quest towards reproducible diagnosis and personalised medicine.

Appendix A

Applications of HoVer-Net

In addition to the experiments that we performed in Chapter 3, we also applied our proposed network for simultaneous segmentation and classification of nuclei to two additional datasets:

- PanNuke dataset [52, 53]
- Multi-Organ Nuclei Segmentation and Classification (MoNuSAC) challenge dataset [147]

PanNuke is now the largest known dataset for nuclear segmentation and classification that ranges across many tissue types and hence is an appropriate additional benchmark of our proposed algorithm. A dataset trained on PanNuke will likely generalise well to new data and therefore can be effectively used for nuclei-based downstream analysis. The MoNuSAC dataset was supplied as part of an international medical imaging contest and therefore the performance of our algorithm compared to other participants' results provides further indication of the ability for our model to segment and classify nuclei. The MoNuSAC dataset[1] was curated to help better understand the tumour microenvironment (TME) and its role in cancer development. For instance, the spatial arrangement of tumour infiltrating lymphocytes (TILs) is associated with clinical outcome in several cancers and tumour associated macrophages (TAMs) influence multiple processes such as blood vessel formation, cell proliferation and antigen presentation in various tumours.

[1] `https://monusac-2020.grand-challenge.org`

A.1 The Datasets

A.1.1 PanNuke 2019

The PanNuke dataset consists of a total of 481 visual fields of H&E stained tissue containing 205,343 annotated nuclei, that have been semi-automatically annotated and quality-controlled by clinical pathologists. The dataset contains nuclei from 19 different tissue types, where image regions were selected with minimal selection bias to better reflect the true data distribution of histology images. Overall, the classes labelled were: neoplastic, inflammatory, connective, epithelial and dead. To generate the dataset, the nuclei were annotated in a two stage process. First a semi-automatic nuclear detection and classification algorithm was used to label nuclei. Then, after several rounds of verification, masks were generated from the detected points [73]. This semi-automatic labeling strategy enabled the creation of a huge dataset, while barely compromising on annotation accuracy. The dataset includes pre-extracted patches of size 256×256 that are split into 3 training, validation and testing folds for a fair model comparison. These folds are selected randomly, but special attention is given to ensure that all tissue types are similarly represented between folds. This is particularly important for minority classes. Example image patches from the PanNuke dataset can be viewed in Figure A.1

A.1.2 MoNuSAC 2020

The MoNuSAC training dataset consists of 31,411 hand-annotated nuclei containing 14,539 epithelial cells, 15,654 lymphocytes, 587 macrophages and 631 neutrophils. Other nuclei, such as fibroblasts and endothelial cells were considered as background and therefore *all* nuclei weren't labelled. It is evident that there is a significant class imbalance in the dataset, which reflects how often the nuclei occur in the tissue. These nuclei were extracted from lung, prostate, breast and kidney H&E tissue sections from 45 patients, which were scanned at 31 hospitals and downloaded from the TCGA database. This enables the subsequent automatic morphological and spatial analysis of the TME, which may help us better understand the role of immune cells in cancer progression. The test set was provided to the challenge participants, but the GT was held back by the organisers. Figure B.1 displays some example image regions from the MoNuSAC dataset, where the colour of the boundary denotes the class of the nuclei.

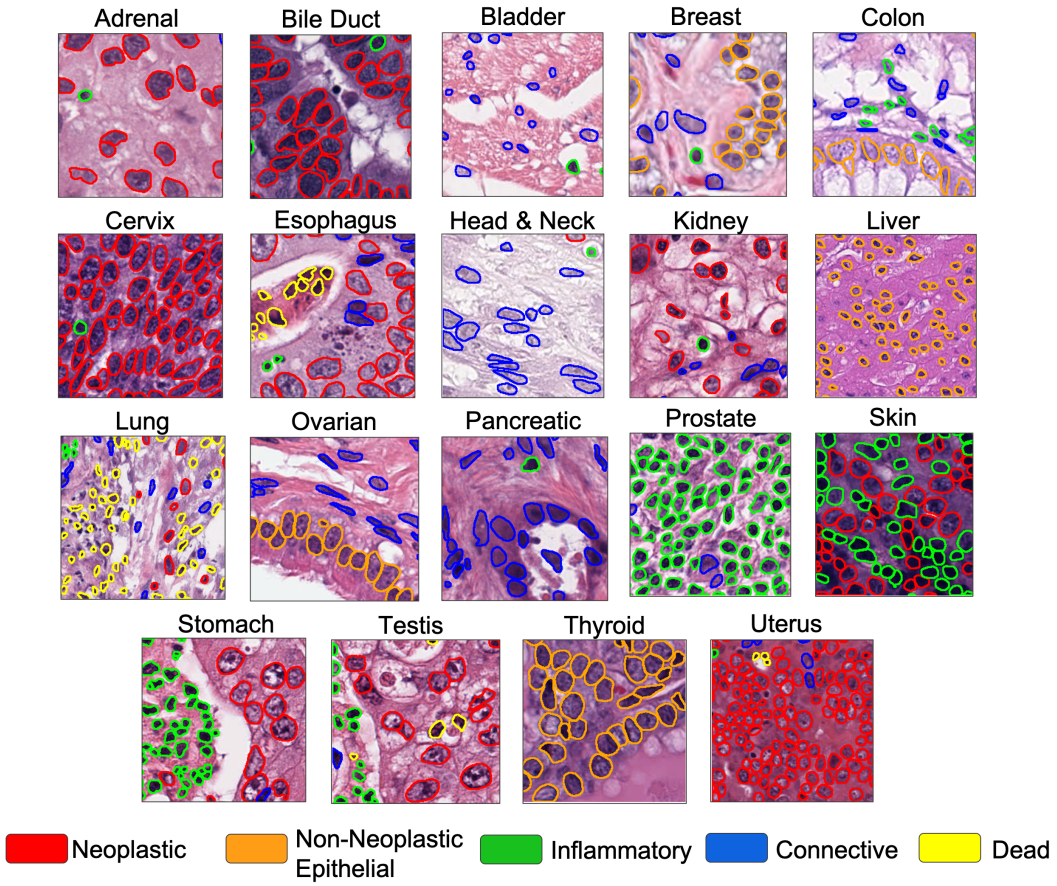


Figure A.1: Example image patches from the PanNuke dataset. The boundary colour denotes the category of each nucleus. Images are taken from [53].

A.2 Experiments and Results

A.2.1 Evaluation Metrics

In Section 3.3.1, we introduced the Panoptic Quality (PQ) as a strong measure to quantify the instance segmentation performance. Then, we proposed a new classification measure in Section 3.3.2 that enabled cross-comparison with detection methods and also on datasets where only the detection point is provided (e.g CRCHisto). However, when only comparing segmentation approaches on a dataset where the segmentation masks for each class are available, it makes sense to extend PQ to a multi-class setting. The multi-class PQ is calculated independently for each class and then the results for each class are averaged to yield the overall result. Due to averaging over the classes, this measure is also insensitive to class imbalance. When calculating mPQ, we skip the PQ calculation for a given class in an image if

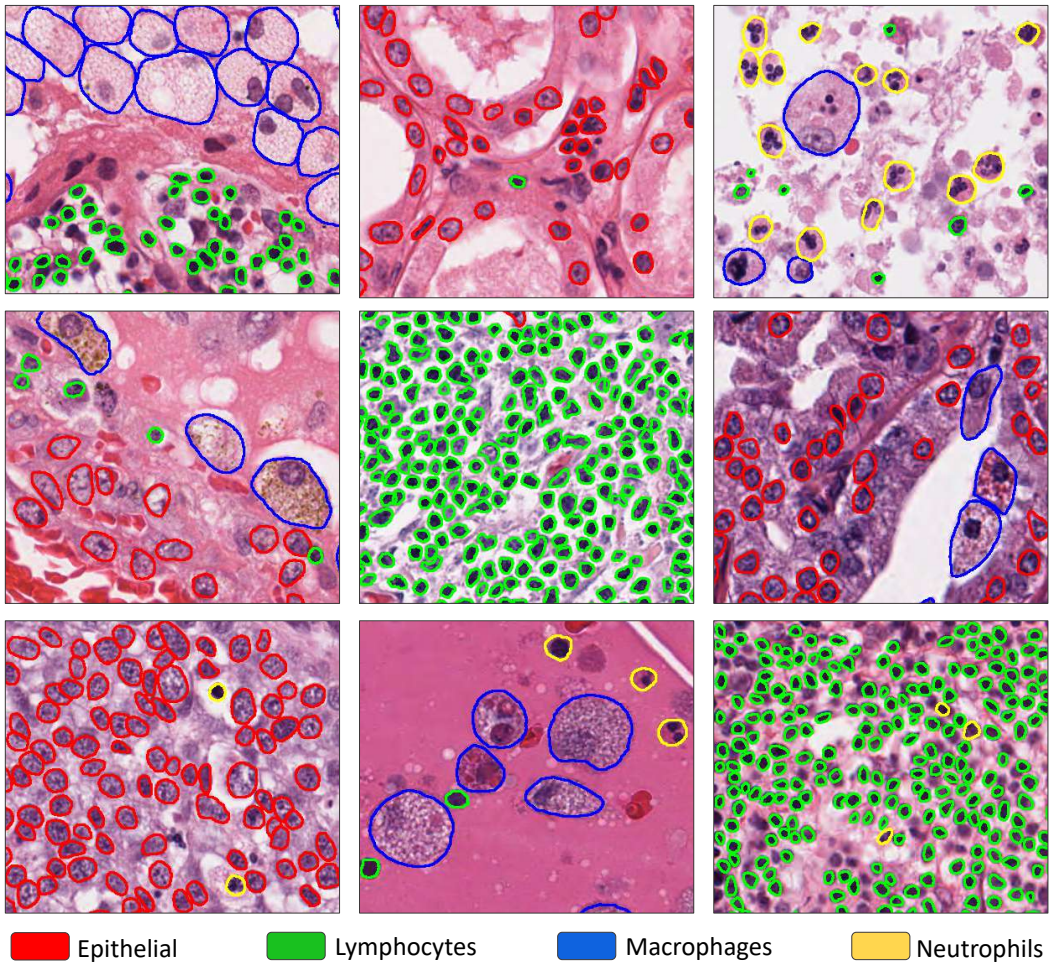


Figure A.2: Example image patches from the MoNuSAC dataset. The boundary colour denotes the category of each nucleus.

its corresponding ground truth mask is empty. We denote binary PQ as bPQ and multi-class PQ as mPQ.

A.2.2 PanNuke Results

We trained HoVer-Net on the 3 folds in PanNuke and report the mPQ and bPQ scores. We are most interested in mPQ as it determines the overall ability for each model to simultaneously segment and classify nuclei into the 5 classes. Similar to our comparative analysis reported in Table 3.5, we include the results of Mask-RCNN [65], DIST [113] and Micro-Net [119] to effectively quantify the performance of our model compared to recent state-of-the-art approaches. In Table A.1, we observe that HoVer-Net achieves the best bPQ and mPQ score across the 19 different tissue

Table A.1: Average mPQ and bPQ across three dataset splits. We also provide the standard deviation (SD) across these splits in the final row.

	DIST		**Mask-RCNN**		**Micro-Net**		**HoVer-Net**	
	mPQ	**bPQ**	**mPQ**	**bPQ**	**mPQ**	**bPQ**	**mPQ**	**bPQ**
Adrenal Gland	0.3442	0.5603	0.3470	0.5546	0.4153	0.6440	0.4812	0.6962
Bile Duct	0.3614	0.5384	0.3536	0.5567	0.4124	0.6232	0.4714	0.6696
Bladder	0.4463	0.5625	0.5065	0.6049	0.5357	0.6488	0.5792	0.7031
Breast	0.3790	0.5466	0.3882	0.5574	0.4407	0.6029	0.4902	0.6470
Cervix	0.3371	0.5309	0.3402	0.5483	0.3795	0.6101	0.4438	0.6652
Colon	0.2989	0.4508	0.3122	0.4603	0.3414	0.4972	0.4095	0.5575
Esophagus	0.3942	0.5295	0.4311	0.5691	0.4668	0.6011	0.5085	0.6427
Head & Neck	0.3177	0.4764	0.3946	0.5457	0.3668	0.5242	0.4530	0.6331
Kidney	0.3339	0.5727	0.3553	0.5092	0.4165	0.6321	0.4424	0.6836
Liver	0.3441	0.5818	0.4103	0.6085	0.4365	0.6666	0.4974	0.7248
Lung	0.2809	0.4978	0.3182	0.5134	0.3370	0.5588	0.4004	0.6302
Ovarian	0.3789	0.5289	0.4337	0.5784	0.4387	0.6013	0.4863	0.6309
Pancreatic	0.3395	0.5343	0.3624	0.5460	0.4041	0.6074	0.4600	0.6491
Prostate	0.3810	0.5442	0.3959	0.5789	0.4341	0.6049	0.5101	0.6615
Skin	0.2627	0.5080	0.2665	0.5021	0.3223	0.5817	0.3429	0.6234
Stomach	0.3369	0.5553	0.3684	0.5976	0.3872	0.6293	0.4726	0.6886
Testis	0.3278	0.5548	0.3512	0.5420	0.4088	0.6300	0.4754	0.6890
Thyroid	0.2574	0.5596	0.3037	0.5712	0.3712	0.6555	0.4315	0.6983
Uterus	0.3487	0.5246	0.3683	0.5589	0.3965	0.5821	0.4393	0.6393
Average across tissues	**0.3406**	**0.5346**	**0.3688**	**0.5528**	**0.4059**	**0.6053**	**0.4629**	**0.6596**
SD across splits	0.0156	0.00975	0.00465	0.00762	0.00816	0.00499	0.00758	0.00364

types present in the dataset and therefore can effectively be used to successfully segment and differentiate between different types of nuclei. Also, our model has the smallest standard deviation across the 3 folds, indicating that it consistently gave a strong performance. In Table A.2 we show the PQ for each class in the dataset, where we can see that our proposed algorithm obtains the best PQ score for each

class. Notably, dead nuclei obtain a low PQ for all models because these nuclei are typically very small and therefore achieving an IoU>0.5 (PQ criterion for a true positive) is difficult. Also, dead nuclei are under-represented in PanNuke, where they make up around 1.5% of the total nuclei in the dataset. Therefore, this class imbalance adds to the difficulty in successful dead nuclei segmentation. Despite this, it is evident that the HoVer-Net score for PQ_d is significantly better than other models because the addition of the dice loss term at the output of the NC branch enables it to perform well when faced with unbalanced classes.

Table A.2: Average PQ for each type of nucleus on the PanNuke dataset. PQ_n, PQ_e, PQ_i, PQ_c and PQ_d denote the panoptic quality for the neoplastic, non-neoplastic epithelial, inflammatory, connective tissue and dead cell classes respectively.

	$\mathbf{PQ}_n$	$\mathbf{PQ}_e$	$\mathbf{PQ}_i$	$\mathbf{PQ}_c$	$\mathbf{PQ}_d$
HoVer-Net	**0.551**	**0.491**	**0.417**	**0.388**	**0.139**
Micro-Net	0.504	0.442	0.333	0.334	0.051
Mask-RCNN	0.472	0.403	0.290	0.300	0.069
DIST	0.439	0.290	0.343	0.275	0.000

A.2.3 MoNuSAC Results

We trained our proposed model for nuclear segmentation and classification on each of the 5 folds and report the performance on each fold in Table A.3. We observe that our model performs best for neutrophils and finds it challenging to segment macrophages. Neutrophils have a clear multi-lobed structure, which allows them to be fairly easily distinguished from other nuclei types. However, macrophages often have an indistinct cell boundary and can significantly vary in their appearance. Therefore, the lower performance for this cell type is expected. In Table A.4 we display the comparative results with other competitors of the MoNuSAC contest. We observe that HoVer-Net achieves the best multi-class PQ score[2], where it obtains a score that is 5.6% higher than second place and 20.4% higher than third place. We show some example results on the MoNuSAC test set in Figure A.3, where we observe that on the whole, our algorithm is able to successfully segment and classify the different nuclei. This further signifies that HoVer-Net is the current state-of-the-art approach for nuclear segmentation and classification.

[2] `https://monusac-2020.grand-challenge.org/Results`

Table A.3: Result on the MoNuSAC dataset for each fold using HoVer-Net. PQ_e, PQ_l, PQ_m and PQ_n denote the PQ panoptic quality for the epithelial, lymphocyte, macrophage and neutrophil classes respectively.

	bPQ	**mPQ**	**PQ_e**	**PQ_l**	**PQ_m**	**PQ_n**
Fold One	0.653	0.460	0.455	0.449	0.406	0.528
Fold Two	0.657	0.534	0.449	0.410	0.577	0.706
Fold Three	0.618	0.466	0.446	0.467	0.467	0.482
Fold Four	0.641	0.471	0.507	0.414	0.464	0.580
Fold Five	0.576	0.462	0.395	0.424	0.413	0.510
Average	0.629	0.478	0.450	0.433	0.465	0.561

Table A.4: Final results of the MoNuSAC contest. *Our submission using HoVer-Net.

Team Name	**mPQ**
TIA-Lab*	**0.6119**
SJTU_426	0.5793
IVG	0.5084
LSL000UD	0.4969
Sharif HooshPardaz	0.4808
xperience.ai	0.4490
TeamTiger	0.4264
Amirreza Mahbod	0.3890
DeepBlueAI	0.3365
Debut_Kele	0.2630
the_great_backpropagator	0.1838
StevenSmiley	0.1659
NUKMLMA	0.1494

A.2.4 Implementation and Training Details

For both above experiments we implemented our framework with the open source software library TensorFlow version 1.8.0 [7] on a workstation equipped with two NVIDIA GeForce 1080 Ti GPUs. PanNuke contains pre-extracted patches of size 256×256 and therefore the input size to our network is slightly smaller than what we originally used in Section 3. When experimenting with PanNuke, we initialised the model with pre-trained weights on the ImageNet dataset [41], trained only the decoders for the first 50 epochs, and then fine-tuned all layers for another 50 epochs. Specifically, we used a batch size of 8 and 4 on each GPU for stage one and two respectively. For MoNuSAC we initialised our model with weights trained on the

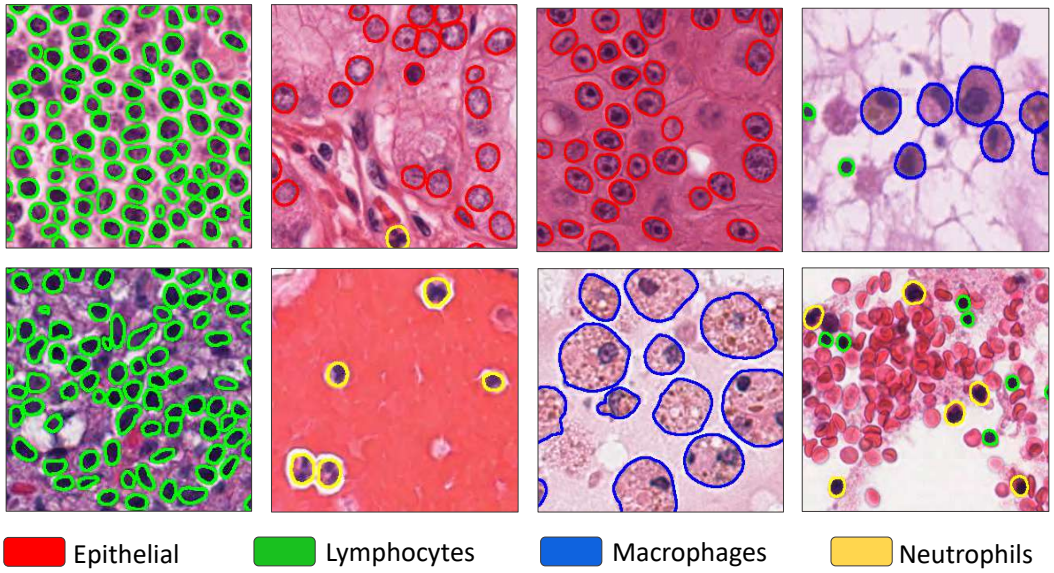


Figure A.3: Visual results on the MoNuSAC dataset.

PanNuke dataset and similarly used an input size of 256×256. Then, we trained only the 1×1 convolutions at the end of each decoder for 12 epochs and then fine-tuned all layers for a further 50 epochs. For both experiments, we used Adam optimisation with an initial learning rate of 10^{-4}, which we then decreased during training. The network was trained with an RGB input, normalised between 0 and 1.

A.3 Discussion and Conclusion

This appendix presented two further applications of our proposed network for simultaneous segmentation and classification of nuclei in histology images. We first applied our algorithm to the PanNuke dataset, which is the largest known multi-tissue dataset for nuclear segmentation and classification and demonstrate that our algorithm is the best performing model compared to recent state-of-the-art approaches. We then applied our model to the MoNuSAC challenge dataset, where we achieved the best performance out of 13 teams. This appendix provided further evidence that HoVer-Net is the current state-of-the-art algorithm for the segmentation and classification of nuclei and enables accurate downstream exploration of nuclear features in the TME.

Appendix B

Exploiting Rotational Symmetry: Additional Experiments and Notation

B.1 Verification of Rotation-Equivariant Approaches

In order to verify our self implemented approaches in Chapter 5, we report the performance of each rotation-equivariant model on the rotated MNIST dataset [92] in Table B.2, which is typically used for performance benchmarking in this domain. The rotated MNIST dataset is a dataset of 70,000 greyscale handwritten digits from 0-9 of size 28×28 pixels, which have been randomly rotated by an angle between 0 and 359°. Therefore, this task requires the model to recognise digits regardless of their orientation. In particular, we report the performance of a conventional CNN, H-Nets [159], standard G-CNNs [32, 21, 90], VF-CNNs [110] and steerable G-CNNs [156]. This was primarily to ensure that we were able to achieve a comparable performance with the reported results in the original papers. In our experiments all CNNs have the same base-level architecture, where we ensured that the models had the same number of layers, the same filter size and a similar number of parameters. Therefore our experiments are not only used for verification, but also to perform a fair head-to-head comparison between models. To maintain a similar number of parameters, we followed the same strategy as described in Section 5.4.4. In line with our experiments in the paper, for H-Net we apply spatial max-pooling based on the magnitudes, as opposed to average-pooling, which is used in the original paper.

We observe that all rotation-equivariant CNNs achieve a greater performance than the conventional CNN, where the best performance is achieved by the C_{12} steer-

Figure B.1: Example images from the MNIST dataset. These images are then rotated by an angle between 0 and 359° to obtain the rotated MNIST dataset.

Table B.1: Verification of baseline models on the rotated MNIST dataset [92]. The superscript associated with H-Net denotes the maximum frequency used.

Method	**Group**	**Parameters**	**Error**
CNN	$\{e\}$	$416K$	2.001
H-Net^1 [159]	$\mathsf{SO}(2)$	$418K$	1.371
H-Net^2 [159]	$\mathsf{SO}(2)$	$414K$	1.352
G-CNN [32]	C_4	$413K$	0.976
G-CNN [21, 90]	C_8	$407K$	0.962
G-CNN [21, 90]	C_{12}	$411K$	0.940
VF-CNN [110]	C_8	$418K$	1.202
VF-CNN [110]	C_{12}	$418K$	1.172
Steerable G-CNN [156]	C_8	$416K$	0.820
Steerable G-CNN [156]	C_{12}	$424K$	**0.809**

able G-CNN. Interestingly, we observe a significant boost in performance for our C_4 G-CNN and H-Net implementations, compared to the originally published results. These models have the same number of layers as the original implementations, but are wider to ensure a similar number of parameters between competing models. Note, we also add 2 1×1 convolutions after obtaining the invariant map (after G-pooling or computing the magnitude of the complex feature maps), which may have also contributed to the increase in performance. If we use the same architecture used by Weiler *et al.* for the C_{12} steerable G-CNN, then we obtain an error of 0.709, which is very close to the original result. However, this implementation uses around $3.3M$ parameters, which is nearly 8× the amount that we use in our comparative experiments in Table B.2.

B.2 Summary of Mathematical Notation in Chapter 5

Table B.2: Description of mathematical symbols.

Symbol	Description
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
$\mathbb{Z}$	Set of integers
$\mathcal{F}$	Real vector space of functions $\mathbb{C} \to \mathbb{R}$
$\mathcal{F}_\mathbb{C}$	Complex vector space of functions $\mathbb{C} \to \mathbb{C}$
Re	Real part of complex number
$\mathsf{E}(2)$	Euclidean group
$\mathsf{SE}(2)$	Special euclidean group (no reflections)
$\mathsf{SO}(2)$	Special orthogonal group (no reflections)
$\{e\}$	Trivial group containing only the identity on page 92
n	A positive integer, fixed throughout this paper
D_n	Dihedral group of n rotations of $\mathbb{C}$, fixing 0 and flips
C_n	Cyclic group of n rotations of $\mathbb{C}$, fixing 0
C'_n	$\{2\pi s/n \mid 0 \leq s < n\}$ group law is addition mod 2π
$\mathcal{G}$	An arbitrary group
G	Group as defined in Subsection 5.2.3
r	radius in polar coordinates
ψ	a filter
λ, β, θ	usually elements of C'_n, sometimes arbitrary angles
R_k	Radial profile of atomic steerable filters

Bibliography

[1] Cancer research uk- bowel cancer statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer. Accessed: 15-04-2020.

[2] Cancer research uk- cancer types. https://www.cancerresearchuk.org/what-is-cancer/how-cancer-starts/types-of-cancer. Accessed: 15-04-2020.

[3] Cancer research uk- lung cancer statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer. Accessed: 15-04-2020.

[4] Meeting pathology demand - histopathology workforce census 2017/18. https://www.rcpath.org/discover-pathology/news/college-report-finds-severe-staff-shortages-across-services-vital-to-cancer-diagnosis.html. Accessed: 25-03-2020.

[5] National cancer institute- cancer classification. https://training.seer.cancer.gov/disease/categories/classification.html. Accessed: 15-04-2020.

[6] World health organisation fact sheet. https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed: 19-03-2020.

[7] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.

[8] Abhinav Agarwalla, Muhammad Shaban, and Nasir M Rajpoot. Representation-aggregation networks for segmentation of multi-gigapixel histology images. *arXiv preprint arXiv:1707.08814*, 2017.

[9] Saad Ullah Akram, Talha Qaiser, Simon Graham, Juho Kannala, Janne Heikkilä, and Nasir Rajpoot. Leveraging unlabeled whole-slide-images for mitosis detection. In *Computational Pathology and Ophthalmic Medical Image Analysis*, pages 69–77. Springer, 2018.

[10] Shadi Albarqouni, Christoph Baur, Felix Achilles, Vasileios Belagiannis, Stefanie Demirci, and Nassir Navab. Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE transactions on medical imaging*, 35(5):1313–1321, 2016.

[11] Sahirzeeshan Ali and Anant Madabhushi. 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, 2012.

[12] Najah Alsubaie, Korsuk Sirinukunwattana, Shan E Ahmed Raza, David Snead, and Nasir Rajpoot. A bottom-up approach for tumour differentiation in whole slide images of lung adenocarcinoma. In *Medical Imaging 2018: Digital Pathology*, volume 10581, page 105810E. International Society for Optics and Photonics, 2018.

[13] Anurag Arnab, Ondrej Miksik, and Philip H. S. Torr. On the robustness of semantic segmentation models to adversarial attacks. *CoRR*, abs/1711.09856, 2017.

[14] Eirini Arvaniti, Kim S Fricker, Michael Moret, Niels Rupp, Thomas Hermanns, Christian Fankhauser, Norbert Wey, Peter J Wild, Jan H Rueschoff, and Manfred Claassen. Automated gleason grading of prostate cancer tissue microarrays via deep learning. *Scientific reports*, 8(1):1–11, 2018.

[15] Ruqayya Awan, Korsuk Sirinukunwattana, David Epstein, Samuel Jefferyes, Uvais Qidwai, Zia Aftab, Imaad Mujeeb, David Snead, and Nasir Rajpoot. Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images. *Scientific reports*, 7(1):16852, 2017.

[16] Aharon Azulay and Yair Weiss. Why do deep convolutional networks generalize so poorly to small image transformations? *arXiv preprint arXiv:1805.12177*, 2018.

[17] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[18] Peter Bankhead, Maurice B Loughrey, José A Fernández, Yvonne Dombrowski, Darragh G McArt, Philip D Dunne, Stephen McQuaid, Ronan T Gray, Liam J Murray, Helen G Coleman, et al. Qupath: Open source software for digital pathology image analysis. *Scientific reports*, 7(1):16878, 2017.

[19] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017.

[20] Babak Ehteshami Bejnordi, Guido Zuidhof, Maschenka Balkenhol, Meyke Hermsen, Peter Bult, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, and Jeroen van der Laak. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. *Journal of Medical Imaging*, 4(4):044504, 2017.

[21] Erik J Bekkers, Maxime W Lafarge, Mitko Veta, Koen AJ Eppenhof, Josien PW Pluim, and Remco Duits. Roto-translation covariant convolutional networks for medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 440–448. Springer, 2018.

[22] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[23] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019.

[24] Anne E Carpenter, Thouis R Jones, Michael R Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A Guertin, Joo Han Chang, Robert A Lindquist, Jason Moffat, et al. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10):R100, 2006.

[25] Hao Chen, Qi Dou, Xi Wang, Jing Qin, Pheng-Ann Heng, et al. Mitosis detection in breast cancer histology images via deep cascaded networks. In *AAAI*, pages 1160–1166, 2016.

[26] Hao Chen, Xiaojuan Qi, Lequan Yu, Qi Dou, Jing Qin, and Pheng-Ann Heng. Dcan: Deep contour-aware networks for object instance segmentation from histology images. *Medical image analysis*, 36:135–146, 2017.

[27] Hao Chen, Xiaojuan Qi, Lequan Yu, and Pheng-Ann Heng. Dcan: deep contour-aware networks for accurate gland segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2487–2496, 2016.

[28] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. 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, 2018.

[29] Jierong Cheng, Jagath C Rajapakse, et al. Segmentation of clustered nuclei with shape markers and marking function. *IEEE Transactions on Biomedical Engineering*, 56(3):741–748, 2009.

[30] Xiuyuan Cheng, Qiang Qiu, Robert Calderbank, and Guillermo Sapiro. Rotdcf: Decomposition of convolutional filters for rotation-equivariant deep networks. In *International Conference on Learning Representations 2019 (ICLR'19)*, 2019.

[31] Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 411–418. Springer, 2013.

[32] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999, 2016.

[33] Taco S Cohen and Max Welling. Steerable cnns. *arXiv preprint arXiv:1612.08498*, 2016.

[34] Carolyn C Compton. Updated protocol for the examination of specimens from patients with carcinomas of the colon and rectum, excluding carcinoid tumors, lymphomas, sarcomas, and tumors of the vermiform appendix: a basis for checklists. *Archives of pathology & laboratory medicine*, 124(7):1016–1025, 2000.

[35] Lee AD Cooper, Alexis B Carter, Alton B Farris, Fusheng Wang, Jun Kong, David A Gutman, Patrick Widener, Tony C Pan, Sharath R Cholleti, Ashish

Sharma, et al. Digital pathology: Data-intensive frontier in medical imaging. *Proceedings of the IEEE*, 100(4):991–1003, 2012.

[36] Germán Corredor, Xiangxue Wang, Yu Zhou, Cheng Lu, Pingfu Fu, Konstantinos Syrigos, David L Rimm, Michael Yang, Eduardo Romero, Kurt A Schalper, et al. Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non–small cell lung cancer. *Clinical Cancer Research*, 25(5):1526–1534, 2019.

[37] Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L Moreira, Narges Razavian, and Aristotelis Tsirigos. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nature medicine*, 24(10):1559–1567, 2018.

[38] Simon S Cross, Samar Betmouni, Julian L Burton, Asha K Dubé, Kenneth M Feeley, Miles R Holbrook, Robert J Landers, Phillip B Lumb, and Timothy J Stephenson. What levels of agreement can be expected between histopathologists assigning cases to discrete nominal categories? a study of the diagnosis of hyperplastic and adenomatous colorectal polyps. *Modern Pathology*, 13(9):941–944, 2000.

[39] Yuxin Cui, Guiying Zhang, Zhonghao Liu, Zheng Xiong, and Jianjun Hu. A deep learning algorithm for one-step contour aware nuclei segmentation of histopathological images. *arXiv preprint arXiv:1803.02786*, 2018.

[40] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017.

[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[42] SL Edwards, C Roberts, ME McKean, JS Cockburn, RR Jeffrey, and KM Kerr. Preoperative histological classification of primary lung cancer: accuracy of diagnosis and use of the non-small cell category. *Journal of Clinical Pathology*, 53(7):537–540, 2000.

[43] Joann G Elmore, Gary M Longton, Patricia A Carney, Berta M Geller, Tracy Onega, Anna NA Tosteson, Heidi D Nelson, Margaret S Pepe, Kimberly H Allison, Stuart J Schnitt, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *Jama*, 313(11):1122–1132, 2015.

[44] Christopher W Elston and Ian O Ellis. Pathological prognostic factors in breast cancer. i. the value of histological grade in breast cancer: experience from a large study with long-term follow-up. cw elston & io ellis. histopathology 1991; 19; 403–410: Author commentary. *Histopathology*, 41(3a):151–151, 2002.

[45] Jonathan I Epstein. An update of the gleason grading system. *The Journal of urology*, 183(2):433–440, 2010.

[46] Matthew Fleming, Sreelakshmi Ravula, Sergei F Tatishchev, and Hanlin L Wang. Colorectal carcinoma: pathologic aspects. *Journal of gastrointestinal oncology*, 3(3):153, 2012.

[47] William T. Freeman and Edward H Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (9):891–906, 1991.

[48] Wolf Herman Fridman, Franck Pagès, Catherine Sautes-Fridman, and Jérôme Galon. The immune contexture in human tumours: impact on clinical outcome. *Nature Reviews Cancer*, 12(4):298–306, 2012.

[49] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

[50] Yarin Gal. *Uncertainty in Deep Learning*. PhD thesis, University of Cambridge, 2016.

[51] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059, 2016.

[52] Jevgenij Gamper, Navid Alemi Koohbanani, Ksenija Benet, Ali Khuram, and Nasir Rajpoot. Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In *European Congress on Digital Pathology*, pages 11–19. Springer, 2019.

[53] Jevgenij Gamper, Navid Alemi Koohbanani, Simon Graham, Mostafa Jahanifar, Ksenija Benet, Syed Ali Khurram, Ayesha Azam, Katherine Hewitt, and Nasir Rajpoot. Pannuke dataset extension, insights and baselines. *arXiv preprint arXiv:2003.10778*, 2020.

[54] Floyd H Gilles, C Jane Tavaré, E Becker Laurence, Peter C Burger, Allan J Yates, Ian F Pollack, and Jonathan L Finlay. Pathologist interobserver variability of histologic features in childhood brain tumors: results from the ccg-945 study. *Pediatric and Developmental Pathology*, 11(2):108–117, 2008.

[55] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

[56] Adam Goode, Benjamin Gilbert, Jan Harkes, Drazen Jukic, and Mahadev Satyanarayanan. Openslide: A vendor-neutral software foundation for digital pathology. *Journal of pathology informatics*, 4, 2013.

[57] Simon Graham, Hao Chen, Jevgenij Gamper, Qi Dou, Pheng-Ann Heng, David Snead, Yee Wah Tsang, and Nasir Rajpoot. Mild-net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical image analysis*, 52:199–211, 2019.

[58] Simon Graham, David Epstein, and Nasir Rajpoot. Rota-net: Rotation equivariant network for simultaneous gland and lumen segmentation in colon histology images. In *European Congress on Digital Pathology*, pages 109–116. Springer, 2019.

[59] Simon Graham and Nasir M Rajpoot. Sams-net: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 590–594. IEEE, 2018.

[60] Simon Graham, Muhammad Shaban, Talha Qaiser, Syed Ali Khurram, and Nasir Rajpoot. Classification of lung cancer histology images using patch-level summary statistics. In *Medical Imaging 2018: Digital Pathology*, volume 10581, page 1058119. International Society for Optics and Photonics, 2018.

[61] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58:101563, 2019.

[62] Anna Gummeson, Ida Arvidsson, Mattias Ohlsson, Niels Christian Overgaard, Agnieszka Krzyzanowska, Anders Heyden, Anders Bjartell, and Kalle Aström. Automatic gleason grading of h and e stained microscopic prostate images

using deep convolutional neural networks. In *Medical Imaging 2017: Digital Pathology*, volume 10140, page 101400S. International Society for Optics and Photonics, 2017.

[63] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2:147–171, 2009.

[64] Stanley R Hamilton, Lauri A Aaltonen, et al. *Pathology and genetics of tumours of the digestive system*, volume 48. IARC press Lyon:, 2000.

[65] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[66] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[67] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity Mappings in Deep Residual Networks. *ArXiv e-prints*, page arXiv:1603.05027, March 2016.

[68] Emiel Hoogeboom, Jorn WT Peters, Taco S Cohen, and Max Welling. Hexaconv. *arXiv preprint arXiv:1803.02108*, 2018.

[69] Le Hou, Dimitris Samaras, Tahsin M Kurc, Yi Gao, James E Davis, and Joel H Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2424–2433, 2016.

[70] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. *ArXiv e-prints*, page arXiv:1608.06993, August 2016.

[71] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[72] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[73] Mostafa Jahanifar, Navid Alemi Koohbanani, and Nasir Rajpoot. Nuclick: From clicks in the nuclei to nuclear boundaries. *arXiv preprint arXiv:1909.03253*, 2019.

[74] Sajid Javed, Muhammad Moazam Fraz, David Epstein, David Snead, and Nasir M Rajpoot. Cellular community detection for tissue phenotyping in histology images. In *Computational Pathology and Ophthalmic Medical Image Analysis*, pages 120–129. Springer, 2018.

[75] Deepak Kademani, R Bryan Bell, Shahrokh Bagheri, Eric Holmgren, Eric Dierks, Bryce Potter, and Louis Homer. Prognostic factors in intraoral squamous cell carcinoma: the influence of histologic grade. *Journal of oral and maxillofacial surgery*, 63(11):1599–1605, 2005.

[76] Jakob Nikolas Kather, Lara R Heij, Heike I Grabsch, Loes FS Kooreman, Chiara Loeffler, Amelie Echle, Jeremias Krause, Hannah Sophie Muti, Jan M Niehues, Kai AJ Sommer, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *bioRxiv*, page 833756, 2019.

[77] Jakob Nikolas Kather, Lara R Heij, Heike I Grabsch, Chiara Loeffler, Amelie Echle, Hannah Sophie Muti, Jeremias Krause, Jan M Niehues, Kai AJ Sommer, Peter Bankhead, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nature Cancer*, 1(8):789–799, 2020.

[78] Jakob Nikolas Kather, Alexander T Pearson, Niels Halama, Dirk Jäger, Jeremias Krause, Sven H Loosen, Alexander Marx, Peter Boor, Frank Tacke, Ulf Peter Neumann, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature medicine*, 25(7):1054–1056, 2019.

[79] Wei Ke, Jie Chen, Jianbin Jiao, Guoying Zhao, and Qixiang Ye. Srn: side-output residual network for object symmetry detection in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1068–1076, 2017.

[80] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in neural information processing systems*, pages 5574–5584, 2017.

[81] Adnan Mujahid Khan, Nasir Rajpoot, Darren Treanor, and Derek Magee. A nonlinear mapping approach to stain normalization in digital histopathol-

ogy images using image-specific color deconvolution. *IEEE Transactions on Biomedical Engineering*, 61(6):1729–1738, 2014.

[82] Mina Khoshdeli and Bahram Parvin. Deep leaning models delineates multiple nuclear phenotypes in h&e stained histology sections. *arXiv preprint arXiv:1802.04427*, 2018.

[83] Alexander Kirillov, Kaiming He, Ross B. Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. *CoRR*, abs/1801.00868, 2018.

[84] K Komuta, K Batts, J Jessurun, D Snover, J Garcia-Aguilar, D Rothenberger, and R Madoff. Interobserver variability in the pathological assessment of malignant colorectal polyps. *British journal of surgery*, 91(11):1479–1484, 2004.

[85] Bin Kong, Xin Wang, Zhongyu Li, Qi Song, and Shaoting Zhang. Cancer metastasis detection via spatially structured deep network. In *International Conference on Information Processing in Medical Imaging*, pages 236–248. Springer, 2017.

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

[87] Neeraj Kumar, Ruchika Verma, Deepak Anand, Yanning Zhou, Omer Fahri Onder, Efstratios Tsougenis, Hao Chen, Pheng Ann Heng, Jiahui Li, Zhiqiang Hu, et al. A multi-organ nucleus segmentation challenge. *IEEE transactions on medical imaging*, 2019.

[88] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging*, 36(7):1550–1560, 2017.

[89] Jin Tae Kwak, Stephen M Hewitt, Sheng Xu, Peter A Pinto, and Bradford J Wood. Nucleus detection using gradient orientation information and linear least squares regression. In *Medical Imaging 2015: Digital Pathology*, volume 9420, page 94200N. International Society for Optics and Photonics, 2015.

[90] Maxime W Lafarge, Erik J Bekkers, Josien PW Pluim, Remco Duits, and Mitko Veta. Roto-translation equivariant convolutional networks: Application to histopathology image analysis. *arXiv preprint arXiv:2002.08725*, 2020.

[91] Dmitry Laptev, Nikolay Savinov, Joachim M Buhmann, and Marc Pollefeys. Ti-pooling: transformation-invariant pooling for feature learning in convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 289–297, 2016.

[92] Hugo Larochelle, Dumitru Erhan, Aaron Courville, James Bergstra, and Yoshua Bengio. An empirical evaluation of deep architectures on problems with many factors of variation. In *Proceedings of the 24th international conference on Machine learning*, pages 473–480, 2007.

[93] A. LaTorre, L. Alonso-Nanclares, S. Muelas, J.M. Peña, and J. DeFelipe. Segmentation of neuronal nuclei based on clump splitting and a two-step binarization of images. *Expert Systems with Applications*, 40(16):6521 – 6530, 2013.

[94] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[95] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[96] Hao Li, Asim Kadav, Igor Durdanovic, Hanan Samet, and Hans Peter Graf. Pruning filters for efficient convnets. *arXiv preprint arXiv:1608.08710*, 2016.

[97] Miao Liao, Yu qian Zhao, Xiang hua Li, Pei shan Dai, Xiao wen Xu, Jun kai Zhang, and Bei ji Zou. Automatic segmentation for cell images based on bottleneck detection and ellipse fitting. *Neurocomputing*, 173:615 – 622, 2016.

[98] Huangjing Lin, Hao Chen, Qi Dou, Liansheng Wang, Jing Qin, and Pheng-Ann Heng. Scannet: A fast and dense scanning framework for metastastic breast cancer detection from whole-slide image. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 539–546. IEEE, 2018.

[99] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[100] Jasper Linmans, Jim Winkens, Bastiaan S Veeling, Taco S Cohen, and Max Welling. Sample efficient semantic segmentation using rotation equivariant convolutional networks. *arXiv preprint arXiv:1807.00583*, 2018.

[101] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[102] Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell. Molecular cell biology 4th edition. *National Center for Biotechnology Information, Bookshelf*, 2000.

[103] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[104] Cheng Lu, David Romo-Bucheli, Xiangxue Wang, Andrew Janowczyk, Shridar Ganesan, Hannah Gilmore, David Rimm, and Anant Madabhushi. Nuclear shape and orientation features from h&e images predict survival in early-stage estrogen receptor-positive breast cancers. *Laboratory Investigation*, 98(11):1438, 2018.

[105] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data efficient and weakly supervised computational pathology on whole slide images. *arXiv preprint arXiv:2004.09666*, 2020.

[106] Ming Y Lu, Melissa Zhao, Maha Shady, Jana Lipkova, Tiffany Y Chen, Drew FK Williamson, and Faisal Mahmood. Deep learning-based computational pathology predicts origins for cancers of unknown primary. *arXiv preprint arXiv:2006.13932*, 2020.

[107] Alessandro Lugli, Richard Kirsch, Yoichi Ajioka, Fred Bosman, Gieri Cathomas, Heather Dawson, Hala El Zimaity, Jean-François Fléjou, Tine Plato Hansen, Arndt Hartmann, et al. Recommendations for reporting tumor budding in colorectal cancer based on the international tumor budding consensus conference (itbcc) 2016. *Modern pathology*, 30(9):1299–1311, 2017.

[108] Marc Macenko, Marc Niethammer, James S Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1107–1110. IEEE, 2009.

[109] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical Image Analysis*, 33:170 – 175, 2016. 20th anniversary of the Medical Image Analysis journal (MedIA).

[110] Diego Marcos, Michele Volpi, Nikos Komodakis, and Devis Tuia. Rotation equivariant vector field networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5048–5057, 2017.

[111] Nikita Moshkov, Botond Mathe, Attila Kertesz-Farkas, Reka Hollandi, and Peter Horvath. Test-time augmentation for deep learning-based cell segmentation on microscopy images. *bioRxiv*, page 814962, 2019.

[112] Eric Nalisnick and Padhraic Smyth. Learning priors for invariance. In *International Conference on Artificial Intelligence and Statistics*, pages 366–375, 2018.

[113] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, 2018.

[114] Kien Nguyen, Anil K Jain, and Bikash Sabata. Prostate cancer detection: Fusion of cytological and textural features. *Journal of pathology informatics*, 2, 2011.

[115] Daniel C Paech, Adèle R Weston, Nick Pavlakis, Anthony Gill, Narayan Rajan, Helen Barraclough, Bronwyn Fitzgerald, and Maximiliano Van Kooten. A systematic review of the interobserver variability for histology in the differentiation between squamous and nonsquamous non-small cell lung cancer. *Journal of Thoracic Oncology*, 6(1):55–63, 2011.

[116] Talha Qaiser, Abhik Mukherjee, Chaitanya Reddy Pb, Sai D Munugoti, Vamsi Tallam, Tomi Pitkäaho, Taina Lehtimäki, Thomas Naughton, Matt Berseth, Aníbal Pedraza, et al. Her 2 challenge contest: a detailed assessment of automated her 2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology*, 72(2):227–238, 2018.

[117] Talha Qaiser, Yee-Wah Tsang, David Epstein, and Nasir Rajpoot. Tumor segmentation in whole slide images using persistent homology and deep convolutional features. In *Annual Conference on Medical Image Understanding and Analysis*, pages 320–329. Springer, 2017.

[118] Shan E Ahmed Raza, Linda Cheung, David Epstein, Stella Pelengaris, Michael Khan, and Nasir M Rajpoot. Mimonet: Gland segmentation using multi-input-multi-output convolutional neural network. In *Annual Conference on Medical Image Understanding and Analysis*, pages 698–706. Springer, 2017.

[119] Shan E Ahmed Raza, Linda Cheung, Muhammad Shaban, Simon Graham, David Epstein, Stella Pelengaris, Michael Khan, and Nasir M Rajpoot. Micro-net: A unified model for segmentation of various objects in microscopy images. *Medical image analysis*, 52:160–173, 2019.

[120] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer graphics and applications*, 21(5):34–41, 2001.

[121] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[122] Brian Ruffell and Lisa M Coussens. Macrophages and therapeutic resistance in cancer. *Cancer cell*, 27(4):462–472, 2015.

[123] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[124] Prudence Anne Russell and Gavin Michael Wright. Predominant histologic subtype in lung adenocarcinoma predicts benefit from adjuvant chemotherapy in completely resected patients: discovery of a holy grail? *Annals of translational medicine*, 4(1), 2016.

[125] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems*, pages 3856–3866, 2017.

[126] Manish Sapkota, Xiaoshuang Shi, Fuyong Xing, and Lin Yang. Deep convolutional hashing for low dimensional binary embedding of histopathological images. *IEEE Journal of Biomedical and Health Informatics*, 2018.

[127] Astrid N Scholten, Vincent THBM Smit, Henk Beerman, Wim LJ van Putten, and Carien L Creutzberg. Prognostic significance and interobserver variability of histologic grading systems for endometrial carcinoma. *Cancer*, 100(4):764–772, 2004.

[128] M. Shaban, R. Awan, M. M. Fraz, A. Azam, Y. Tsang, D. Snead, and N. M. Rajpoot. Context-aware convolutional neural network for grading of colorectal cancer histology images. *IEEE Transactions on Medical Imaging*, pages 1–1, 2020.

[129] Muhammad Shaban, Ruqayya Awan, Muhammad Moazam Fraz, Ayesha Azam, Yee-Wah Tsang, David Snead, and Nasir M Rajpoot. Context-aware convolutional neural network for grading of colorectal cancer histology images. *IEEE Transactions on Medical Imaging*, 2020.

[130] Harshita Sharma, Norman Zerbe, Daniel Heim, Stephan Wienert, Hans-Michael Behrens, Olaf Hellwich, and Peter Hufnagl. A multi-resolution approach for combining visual information using nuclei segmentation and classification in histopathological images. In *VISAPP (3)*, pages 37–46, 2015.

[131] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

[132] Xiaoshuang Shi, Fuyong Xing, KaiDi Xu, Yuanpu Xie, Hai Su, and Lin Yang. Supervised graph hashing for histopathology image retrieval and classification. *Medical image analysis*, 42:117–128, 2017.

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

[134] Korsuk Sirinukunwattana, Shan e Ahmed Raza, Yee-Wah Tsang, David RJ Snead, Ian A Cree, and Nasir M Rajpoot. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans. Med. Imaging*, 35(5):1196–1206, 2016.

[135] Korsuk Sirinukunwattana, Josien PW Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J Matuszewski, Elia Bruni, Urko Sanchez, et al. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35:489–502, 2017.

[136] Korsuk Sirinukunwattana, David Snead, David Epstein, Zia Aftab, Imaad Mujeeb, Yee Wah Tsang, Ian Cree, and Nasir Rajpoot. Novel digital signatures of tissue phenotypes for predicting distant metastasis in colorectal cancer. *Scientific reports*, 8(1):13692, 2018.

[137] David RJ Snead, Yee-Wah Tsang, Aisha Meskiri, Peter K Kimani, Richard Crossman, Nasir M Rajpoot, Elaine Blessing, Klaus Chen, Kishore Gopalakrishnan, Paul Matthews, et al. Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology*, 68(7):1063–1072, 2016.

[138] Luisa M Solis, Carmen Behrens, M Gabriela Raso, Heather Y Lin, Humam Kadara, Ping Yuan, Hector Galindo, Ximing Tang, J Jack Lee, Neda Kalhor, et al. Histologic patterns and molecular characteristics of lung adenocarcinoma associated with clinical outcome. *Cancer*, 118(11):2889–2899, 2012.

[139] student (https://math.stackexchange.com/users/20150/student). Every measurable homomorphism from $\mathbb{R}^n$ to $\mathbb{C}^*$ is exponential. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/442980 (version: 2013-07-13).

[140] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[141] Elizabeth R Tang, Andrew M Schreiner, and Bradley B Pua. Advances in lung adenocarcinoma classification: a summary of the new international multidisciplinary classification system (iaslc/ats/ers). *Journal of thoracic disease*, 6(Suppl 5):S489, 2014.

[142] David Tellez, Maschenka Balkenhol, Irene Otte-Höller, Rob van de Loo, Rob Vogels, Peter Bult, Carla Wauters, Willem Vreuls, Suzanne Mol, Nico Karssemeijer, et al. Whole-slide mitosis detection in h&e breast histology using phh3 as a reference to train distilled stain-invariant convolutional networks. *IEEE transactions on medical imaging*, 37(9):2126–2136, 2018.

[143] David Tellez, Geert Litjens, Peter Bandi, Wouter Bulten, John-Melle Bokhorst, Francesco Ciompi, and Jeroen van der Laak. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *arXiv preprint arXiv:1902.06543*, 2019.

[144] William D Travis, Elisabeth Brambilla, Andrew G Nicholson, Yasushi Yatabe, John HM Austin, Mary Beth Beasley, Lucian R Chirieac, Sanja Dacic, Edwina Duhig, Douglas B Flieder, et al. The 2015 world health organization classifica-

tion of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *Journal of thoracic oncology*, 10(9):1243–1260, 2015.

[145] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Transactions on Medical Imaging*, 35(8):1962–1971, Aug 2016.

[146] Bastiaan S Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling. Rotation equivariant cnns for digital pathology. In *International Conference on Medical image computing and computer-assisted intervention*, pages 210–218. Springer, 2018.

[147] Neeraj; Patil Abhijeet; Kurian Nikhil; Rane Swapnil; Verma, Ruchika; Kumar and Amit Sethi. Multi-organ nuclei segmentation and classification challenge 2020. 02 2020.

[148] M Veta, PJ van Diest, R Kornegoor, A Huisman, MA Viergever, and JPW Pluim. Automatic nuclei segmentation in h&e stained breast cancer histopathology images. *PLoS ONE*, 8(7):e70221, 2013.

[149] Mitko Veta, Yujing J Heng, Nikolas Stathonikos, Babak Ehteshami Bejnordi, Francisco Beca, Thomas Wollmann, Karl Rohr, Manan A Shah, Dayong Wang, Mikael Rousson, et al. Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge. *Medical image analysis*, 54:111–121, 2019.

[150] Mitko Veta, Paul J Van Diest, Stefan M Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, Anders BL Larsen, Jacob S Vestergaard, Anders B Dahl, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical image analysis*, 20(1):237–248, 2015.

[151] Quoc Dang Vu, Simon Graham, Minh Nguyen Nhat To, Muhammad Shaban, Talha Qaiser, Navid Alemi Koohbanani, Syed Ali Khurram, Tahsin Kurc, Keyvan Farahani, Tianhao Zhao, et al. Methods for segmentation and classification of digital microscopy tissue images. *arXiv preprint arXiv:1810.13230*, 2018.

[152] Zhili Wan, Tiansheng Yin, Hongwei Chen, and Dewei Li. Surgical treatment of a retroperitoneal benign tumor surrounding important blood vessels by fractionated resection: A case report and review of the literature. *Oncology letters*, 11(5):3259–3264, 2016.

[153] Pin Wang, Xianling Hu, Yongming Li, Qianqian Liu, and Xinjian Zhu. Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Signal Processing*, 122:1–13, 2016.

[154] Mary Kay Washington, Jordan Berlin, Philip Branton, Lawrence J Burgart, David K Carter, Patrick L Fitzgibbons, Kevin Halling, Wendy Frankel, John Jessup, Sanjay Kakar, et al. Protocol for the examination of specimens from patients with primary carcinoma of the colon and rectum. *Archives of pathology & laboratory medicine*, 133(10):1539–1551, 2009.

[155] Maurice Weiler and Gabriele Cesa. General e (2)-equivariant steerable cnns. In *Advances in Neural Information Processing Systems*, pages 14334–14345, 2019.

[156] Maurice Weiler, Fred A Hamprecht, and Martin Storath. Learning steerable filters for rotation equivariant cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 849–858, 2018.

[157] Stephan Wienert, Daniel Heim, Kai Saeger, Albrecht Stenzinger, Michael Beil, Peter Hufnagl, Manfred Dietel, Carsten Denkert, and Frederick Klauschen. Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach. *Scientific reports*, 2:503, 2012.

[158] Daniel Worrall and Max Welling. Deep scale-spaces: Equivariance over scale. In *Advances in Neural Information Processing Systems*, pages 7364–7376, 2019.

[159] Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5028–5037, 2017.

[160] Yan Xu, Yang Li, Mingyuan Liu, Yipei Wang, Maode Lai, I Eric, and Chao Chang. Gland instance segmentation by deep multichannel side supervision. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 496–504. Springer, 2016.

[161] Yan Xu, Yang Li, Yipei Wang, Mingyuan Liu, Yubo Fan, Maode Lai, I Eric, and Chao Chang. Gland instance segmentation using deep multichannel neural networks. *IEEE Transactions on Biomedical Engineering*, 64(12):2901–2912, 2017.

[162] Yukako Yagi. Color standardization and optimization in whole slide imaging. In *Diagnostic pathology*, volume 6, page S15. Springer, 2011.

[163] Lin Yang, Yizhe Zhang, Jianxu Chen, Siyuan Zhang, and Danny Z Chen. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 399–407. Springer, 2017.

[164] Xiaodong Yang, Houqiang Li, and Xiaobo Zhou. 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, 2006.

[165] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[166] Kun-Hsing Yu, Ce Zhang, Gerald J Berry, Russ B Altman, Christopher Ré, Daniel L Rubin, and Michael Snyder. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature communications*, 7:12474, 2016.

[167] Yinyin Yuan, Henrik Failmezger, Oscar M Rueda, H Raza Ali, Stefan Gräf, Suet-Feung Chin, Roland F Schwarz, Christina Curtis, Mark J Dunning, Helen Bardwell, et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Science translational medicine*, 4(157):157ra143–157ra143, 2012.

[168] Yizhe Zhang, Lin Yang, Jianxu Chen, Maridel Fredericksen, David P Hughes, and Danny Z Chen. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 408–416. Springer, 2017.

[169] Yanning Zhou, Omer Fahri Onder, Qi Dou, Efstratios Tsougenis, Hao Chen, and Pheng-Ann Heng. Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation. *arXiv preprint arXiv:1903.05358*, 2019.

[170] Yanzhao Zhou, Qixiang Ye, Qiang Qiu, and Jianbin Jiao. Oriented response networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 519–528, 2017.