Computational Models for Automated Histopathological Assessment of Colorectal Liver Metastasis Progression

Zhaoyang Xu

School of Electronic Engineering and Computer Science

Queen Mary University of London

This dissertation is submitted for the degree of

Doctor of Philosophy

September 2019

I would like to dedicate this thesis to my loving parents.

Declaration

I, Zhaoyang Xu, declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Zhaoyang Xu

September 2019

Details of collaboration and publications:

Papers published while working on this thesis are listed at the end of the thesis. Any publications produced in collaboration with others are clearly mentioned. Any further collaborations are labelled as such within the text.

Acknowledgements

I would like to express my sincere gratitude to all the people who supported me during my PhD study. First of all, I would like to thank my supervisor, Dr. Qianni Zhang, for her constant encouragement and guidance. She has walked me through all the stages of my PhD study. I am deeply grateful to Prof. Ebroul, who has supported me all the time. I would also like to thank Dr. Huiyu Zhou and Dr. Pengwei Hao for their many suggestions and encouragements during these years. Meanwhile, I would like to thank Dr. Carlos Fernández Moro and Dr. Béla Bozóky from Karolinska University Hospital for collecting and annotating the dataset.

I am very grateful to my friends and colleagues at the MultiMedia and Vision (MMV) group for the companionship ad valuable discussions and suggestions in the memorable four years, Dr. Heng Yang, Dr. Shenglan Huang, Dr. Wenxuan Mou, Tingting Xie, Yan Huang, Yibao Sun, Yilong Li, Xingru Huang, Bingqing Guo, Xindi Zhang, Bingqun Liu. I'd like to thank my friend Dr. Shan Luo for the discussion and my roommate Xiaoshuai Zhang for his support in the daily life.

Finally, with my love and gratitude, I would like to dedicate this thesis to my family who always give me the unreserved support and selfishless love. It is my best luck to have them in my life.

Abstract

Histopathology imaging is a type of microscopy imaging commonly used for the micro-level clinical examination of a patient's pathology. Due to the extremely large size of histopathology images, especially whole slide images (WSIs), it is difficult for pathologists to make a quantitative assessment by inspecting the details of a WSI. Hence, a computer-aided system is necessary to provide a subjective and consistent assessment of the WSI for personalised treatment decisions. In this thesis, a deep learning framework for the automatic analysis of whole slide histopathology images is presented for the first time, which aims to address the challenging task of assessing and grading colorectal liver metastasis (CRLM). Quantitative evaluations of a patient's condition with CRLM are conducted through quantifying different tissue components in resected tumorous specimens. This study mimics the visual examination process of human experts, by focusing on three levels of information, the tissue level, cell level and pixel level, to achieve the step by step segmentation of histopathology images.

At the tissue level, patches with category information are utilised to analyse the WSIs. Both classification-based approaches and segmentation-based approaches are investigated to locate the metastasis region and quantify different components of the WSI. For the classification-based method, different factors that might affect the classification accuracy are explored using state-of-the-art deep convolutional neural networks (DCNNs). Furthermore, a novel network is proposed to merge the information from different magnification levels to include contextual information to support the final decision. With the support by the

segmentation-based method, edge information from the image is integrated with the proposed fully convolutional neural network to further enhance the segmentation results.

At the cell level, nuclei related information is examined to tackle the challenge of inadequate annotations. The problem is approached from two aspects: a weakly supervised nuclei detection and classification method is presented to model the nuclei in the CRLM by integrating a traditional image processing method and variational auto-encoder (VAE). A novel nuclei instance segmentation framework is proposed to boost the accuracy of the nuclei detection and segmentation using the idea of transfer learning. Afterwards, a fusion framework is proposed to enhance the tissue level segmentation results by leveraging the statistical and spatial properties of the cells.

At the pixel level, the segmentation problem is tackled by introducing the information from the immunohistochemistry (IHC) stained images. Firstly, two data augmentation approaches, synthesis-based and transfer-based, are proposed to address the problem of insufficient pixel level segmentation. Afterwards, with the paired image and masks having been obtained, an end-to-end model is trained to achieve pixel level segmentation. Secondly, another novel weakly supervised approach based on the generative adversarial network (GAN) is proposed to explore the feasibility of transforming unpaired haematoxylin and eosin (HE) images to IHC stained images. Extensive experiments reveal that the virtually stained images can also be used for pixel level segmentation.

Table of contents

List of figures **xi**

List of tables **xiv**

List of Abbreviations **xv**

1 Introduction **1**

1.1 Background 1

1.1.1 Histopathology images - slides preparation and digitisation 2

1.1.2 Colorectal liver metastasis - background 4

1.1.3 Digital histopathology images 7

1.2 Research objective 8

1.2.1 General tasks for histopathology image analysis 8

1.2.2 Research objective of this study 10

1.3 Challenges 10

1.4 Outline of the thesis 12

2 Literature Review **14**

2.1 Related work on tissue patch classification and segmentation 15

2.1.1 Tissue level overview 15

2.1.2 Tissue level classification Method 18

2.1.3 Tissue level segmentation methods . . . 23
2.1.4 Publicly available datasets on tissue/slide level . . . 25
2.2 Related work on cell level tasks . . . 26
2.2.1 Image pre-processing: image deconvolution and normalisation . . . 27
2.2.2 Nuclei detection . . . 28
2.2.3 Nuclei segmentation . . . 29
2.2.4 Nuclei classification . . . 30
2.2.5 Publicly available datasets for cell level . . . 32
2.3 Towards a pixel level analysis with weakly-supervised methods . . . 32
2.3.1 GAN-based networks for dataset augmentation . . . 33
2.3.2 GAN-based networks for image enhancement . . . 34
2.4 Evaluation . . . 35
2.4.1 Evaluation metrics for classification . . . 36
2.4.2 Evaluation metrics for segmentation . . . 36
2.4.3 Other metrics . . . 37
2.5 Open-source platforms for histopathology images management and analysis 38
2.5.1 Traditional cell level platforms . . . 38
2.5.2 Basic libraries or tools . . . 39
2.5.3 PC-based software . . . 40
2.5.4 Web-based platforms . . . 41

3 Deep Neural Networks for Whole Slide Image Analysis 42
3.1 Overview and problem formulation . . . 43
3.2 DCNNs components and data preparation . . . 45
3.2.1 Components & structures of DCNNs . . . 46
3.2.2 Visualising and understanding the network . . . 50
3.2.3 Dataset preparation and augmentation . . . 51

3.3 Multi-scale context-aware convolutional network (MCCNet) 56
3.3.1 Network structure and implement details 57
3.3.2 Results and discussion 58
3.4 Whole slide image analysis using DCNNs 60
3.4.1 Methods and experiments 61
3.4.2 Results and discussion 62
3.5 From tissue level to pixel level 69
3.5.1 Framework and methodology 69
3.5.2 Experiment details 72
3.5.3 Results and discussion 73
3.6 Summary 74

4 Nuclei Segmentation and Classification for Cell Level Histopathology Image Analysis 77

4.1 Overview 78
4.2 Nuclei detection and classification in liver tissue 81
4.2.1 Preliminaries 81
4.2.2 Image pre-processing and nuclei detection 84
4.2.3 Convolutional variational auto-encoder for nuclei classification . . . 88
4.3 US-net for robust nuclei instance segmentation 94
4.3.1 Background and preliminaries 94
4.3.2 US-net architecture 96
4.3.3 Experiments and evaluation 99
4.3.4 Results and discussion 101
4.4 Multi-scale fusion for semantic segmentation in CRLM 102
4.4.1 Framework and proposed method 102
4.4.2 Dataset preparation 106

4.4.3 Implementation details . . . 107
4.4.4 Results and discussion . . . 108
4.5 Summary . . . 109

5 Weakly Supervised Models for Pixel Level Histopathology Image Analysis 111
5.1 Overview . . . 112
5.1.1 Problem statement . . . 112
5.1.2 Challenges and objective . . . 116
5.1.3 Preliminaries . . . 116
5.2 Augmented training for pixel level segmentation . . . 119
5.2.1 Image synthesis for data augmentation . . . 119
5.2.2 Image transformation for data augmentation . . . 120
5.2.3 Dataset collection and preparation . . . 124
5.2.4 Implement details and results . . . 126
5.2.5 Results and discussion . . . 127
5.3 Conditional CycleGAN for virtual restaining . . . 131
5.3.1 Proposed method . . . 131
5.3.2 Dataset and implementation . . . 135
5.3.3 Results and discussion . . . 137
5.4 Summary . . . 146

6 Conclusion 148
6.1 Findings and Limitations . . . 148
6.2 Future Work . . . 149

References 152

List of figures

1.1 Histopathology specimen processing pipeline 3

1.2 Example of an annotated histopathology region 5

1.3 Illustrations of different invasive front patterns 6

1.4 Pyramid structure of WSIs and components observed on different levels . . 7

2.1 Patch size vs. magnification level 16

2.2 Example of tissue patch with different components from TCGA dataset . . 21

3.1 From whole slide images to patches 44

3.2 Comparison between repetitive texture (a) and orderless texture (b) 44

3.3 Understand the WSI from the "Texton" point of view 45

3.4 Mean pooling and max pooling 47

3.5 Network structures of Alexnet, VGG-16 and Resnet-34 48

3.6 First layer feature maps of models trained at different magnification levels. . 50

3.7 High-level feature maps and their corresponding features 51

3.8 Features of top reactions in the layer5 for different patches. 52

3.9 Example of an annotated ROI from CRLM dataset. 53

3.10 Patches from different magnification levels and different classes 54

3.11 Illustration of the data augmentation methods 56

3.12 Proposed MCCNet network structure 57

3.13 Performances of proposed MCCNet on the ROIs 58
3.14 Comparison of F1-scores on different classes 60
3.15 Visualisation of feature vectors from the DCNN on 2d space with t-SNE . . 62
3.16 Whole slide results with model Resnet-34 64
3.17 Examples of results on WSIs with model Resnet-34 and patch size 448 . . . 65
3.18 "Pseudo ground truth" for whole slide segmentation evaluation. 67
3.19 Hierarchical labelling system for CRLM dataset 68
3.20 Proposed framework for segmentation-based segmentation 70
3.21 Simple data synthesis strategy 72
3.22 Examples of the segmentation results. 76

4.1 Cell structure and example of the nuclei 78
4.2 Cell level problem statement and task overview 79
4.3 Illustration of auto-encoder and stacked auto-encoder 82
4.4 Direct graphical model 83
4.5 Nuclei region proposal pipeline. 85
4.6 Example of proposed nuclei regions 88
4.7 Different types of cells in hepatocyte tissue 89
4.8 Network structures of convolutional auto-encoder and variational auto-encoder. 90
4.9 Input image patches (left) and the reconstructed image patches by CAE (right) 91
4.10 Features of the nuclei regions on transformed space. 92
4.11 Samples that used to extract the centre of the features. 93
4.12 Basic framework of the U-net 95
4.13 Network structure of SSD network 96
4.14 Network structure of proposed US-net with post-processing sub-networks . 97
4.15 Losses and evaluation accuracies for different network structures 100
4.16 Results of the proposed method on the CRLM dataset 101

4.17 An overview of the proposed fusion framework 103
4.18 An illustration of the proposed Tissue Region Growing algorithm 105
4.19 Delauney tree of the nuclei's location 105
4.20 Examples of the segmented nuclei from CRLM dataset 107
4.21 The outcomes of the intermediate steps of tissue region growing algorithm. 108
4.22 Examples of segmentation results 110

5.1 Examples of matched HE and IHC stained images patches 113
5.2 Pixel level problem statement and task overview 114
5.3 Network structure for generative adversarial network 118
5.4 Image synthesis framework for data augmentation 120
5.5 Augmented training framework for pixel level segmentation 121
5.6 Example of synthesised image for augment training 128
5.7 Examples of transformed image for augment training 129
5.8 Examples of segmentation results. 130
5.9 Mapping between two domains and their sub-domains 131
5.10 Overall structure of the proposed cCGAN network 132
5.11 Example of the unpaired training dataset 136
5.12 Proposed cCGAN model results with different parameters settings 138
5.13 Unsuccessful mapping examples of the virtual restaining model 140
5.14 Virtual restaining results on the large patches 141
5.15 Virtual restaining and virtual registration results on the whole slide image . 142
5.16 Example results for image registration and multipixel 144
5.17 Colour deconvolution examples using virtual staining framework 145
5.18 Examples of virtual staining results. 147

List of tables

2.1 Related research using CAMELYON dataset 21
2.2 A summary of the methods employed for BACH contest 22
2.3 Public dataset for tissue-level histopathology image analysis 25
2.4 Public datasets for cell-level histopathology image analysis 31

3.1 Details of the annotated WSIs 54
3.2 Overall quantitative classification accuracy of different models 59
3.3 Classification accuracy on the CRLM dataset 63
3.4 Evaluation accuracy on colorectal lymph node metastases 66
3.5 Segmentation results (PA & MIoU) on the evaluation dataset with different models 74

4.1 Encoder and decoder layout of FCN 90
4.2 Evaluation of the classification precision and recall 93
4.3 Quantitative segmentation result (PA & MIoU) of proposed methods 108

5.1 Comparison of segmentation results (PA & MIoU) among different models 128
5.2 Quantitative evaluation results of the models for virtual restaining 139
5.3 Segmentation results (PA & MIoU) based on virtual restained image 143

List of Abbreviations

AI Artificial Intelligence

CAD Computer-Aided Diagnosis

cCGAN conditional CycleGAN

CK CytoKeratine

CRCLNM Colorectal Lymph Node Metastasis

CRF Random Conditional Field

CRLM Colorectal Liver Metastasis

CVM Conditional Virtual Magnification

DCNN Deep Convolutional Neural Network

FCN Fully Convolutional Neural Network

GAN Generative Adversarial Network

HE haematoxylin and Eosin

IHC ImmunoHistoChemistry

KUH Karolinska University Hospital

mAP mean Average Precision

MCCNet Multi-scale Context-aware Convolutional Network

MIoU Mean Intersection of Union

MRI Magnetic Resonance Imaging

NMF Non-negative Matrix Factorisation

PA Pixel Accuracy

ReLU Rectified Linear Unit

RNN Recurrent Neural Network

ROC Receiver Operating Characteristic

ROI Region Of Interest

SGD Stochastic Gradient Descent

SSD Single Shot Detection

t-SNE t-distributed Stochastic Neighbour Embedding

VR Virtual Registration

WSI Whole Slide Image

CVAE Convolutional Variational Auto-Encoder

VAE Variational Auto-Encoder

Chapter 1

Introduction

The analysis of histology images is an essential approach for biologists and pathologists to study an object's characteristics at a micro level, as the microscopic images of cells and tissues contain valuable information for both scientific study and medical diagnosis [1]. Especially for cancer diagnosis and progress evaluation, histology images (referred to as histopathology images) are considered as the "golden truth" [2]. In this research, the main focus is developing computer-aided histopathology image analysis algorithms for assisting regression assessment for colorectal liver metastases. In this chapter, the background, research objective and challenges of this research are introduced and then the outline of the thesis is provided.

1.1 Background

Histopathology images refer to the microscopic images of animal tissues or cells used for the purpose of clinical pathology diagnosis. In the medical field, the condition of the cells and tissue can provide a reliable interpretation of the corresponding performance of the patients' organs. For biological study, the composition of the tissue or the growth pattern of the cells is an effective reflection of the lower level mechanism. However, these are too small to

be observed directly with the eyes. Only with the help of a microscope, can the object be explicitly inspected after being magnified dozens of times. Therefore, microscopy imaging becomes an indispensable technique for biomedical study.

Most histopathology images are obtained using a light microscope, but there are other kinds of equipment available such as electron or infrared scopes. However, these are not widely used due to the cost or other factors. Currently, there are some problems with the traditional microscope [3]. For example, the slides need to be loaded for the observation whenever they are needed. This is inconvenient and inefficient particularly for slide analysis on a large scales. Thanks to recent developments in advanced microscopy, the images of slides at different magnification levels can be easily digitalised and stored. Subsequently, the slide can be viewed on any computer locally or remotely. Moreover, observers can obtain a better view on computer screens compared to using scopes.

Aside from the imaging methods, how the tissue is acquired also makes a difference. Some slides are made from biopsies, while others are from specimens. With consideration of the staining methods, there are many more types of histopathology images in terms of their appearance. In this study, the main research focus is on light microscope scanned histopathology images from resected tissue after the surgery for colorectal liver metastasis (CRLM) patients.

1.1.1 Histopathology images - slides preparation and digitisation

Unlike the acquisition of other medical images like X-ray and Ultrasound, the acquisition of histopathology images is rather complicated. Typically, there are six steps (as shown in Fig. 1.1) to prepare a slide for digitalization in a pathology lab: fixation, grossing, processing, embedding, sectioning and staining [4]. The purpose of fixation is to prevent the tissues from autolysis, and to preserve them in the same state as when they were alive. Fixation also hardens the specimen to maintain its molecular structure. Grossing is a relatively simple

operation that occurs in the grossing station, where the specimens are labelled, measured and pre-processed. At the processing step, the specimens react with different types of chemical reagents in an automated processor machine. Then, after the embedding, the specimens are solidified on the platform to be readied for sectioning cutting. A thin piece of tissue is placed on the slide, and then the last step is to stain the tissue on the slide, otherwise, it is not possible to observe the slide since the thin piece of tissue would otherwise be almost transparent.

There are a series of dyes such as hematoxylin & eosin (HE) and immunohistochenistry (IHC) that can be used to highlight the different components of the tissue. However, HE staining is the most commonly used and low-cost staining method and is used to highlight the nuclei out of the cytoplasm and reveal the structure of the tissue as well. The whole process takes around 12 to 24 hours before the slide can be sent for digitalisation depending on the size of the specimen. Compared to X-ray and MRI, the acquisition of a histopathology image takes a very long time.

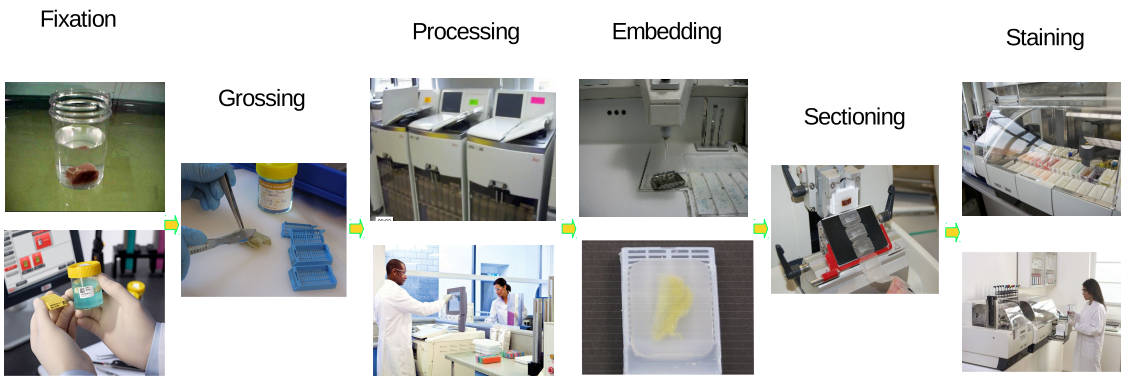


Fig. 1.1 Steps for specimen preparation with figures from Geoffrey Rolls' [1] and Chris Mills' [2] tutorials

[1] Six Steps of Processing a Biopsy in the Histology Lab, from website: *"https://owlcation.com/stem/what-happened-to-that-biopsy-the-doctor-took-from-me/"*, accessed at September, 2017

[2] An Introduction to Specimen Preparation, *"http://www.leicabiosystems.com/pathologyleaders/an-introduction-to-specimen-preparation/"*, accessed at September, 2017

1.1.2 Colorectal liver metastasis - background

Colorectal cancer (bowel cancer, CRC) is one of the most common cancers in the world, where over 1 million patients are diagnosed with colorectal cancer each year [5, 6]. There are four stages of colorectal cancer, the final stage being metastases, which means the cancer has spread to other parts of the body. CRLM is a very serious stage for the patient since for most colorectal cancer patients, the liver is the first target organ for distant metastases. Some patients may develop CRLM along with their original cancer, or even after their original cancer has been treated. The standard care for patients with CRLM is surgical resection which is also the only effective treatment for longer survival or cure. However, according to the distribution of the disease, only about 20% CRLM patients are identified as being suitable for resection [7]. Even though the 5-year survival rate with resected CRLM has increased to 50-60% with the progress made in a multidisciplinary field [8], the overall 5-year survival rate is too low, and this makes the treatment after resection so important.

Before resection surgery, patients have to undertake chemotherapy to control the growth of the tumour tissue and decrease the size of the tumour. Furthermore, preoperative chemotherapy will help 16% patients with initially unresectable lesions to achieve a resectable condition [9]. However, even though preoperative chemotherapy helps the prognosis of patients, it is impossible to eliminate all tumour cells since this will cause damage to the healthy parts of the body as well. Thus, after surgery, the number of residual tumour cells in a sample can be considered as critical information to inform the subsequent treatment strategy. Aside from the histopathology image, other medical images such as from magnetic resonance imaging (MRI) can be used to estimate the tumour size. However, after chemotherapy, the other methods cannot be used to estimate the tumour size as they cannot access the tumour to assess its condition.

After surgery, patients are still subject to cancer related risks. Recurrence is the most significant risk facing CRLM patients after resection surgery, making after-surgery evaluation

and a recovery plan so important. With the resected tissue making it possible to obtain a histopathology image, this provides pathologists with opportunity to perform detailed regression grading. Afterwards, a personalised plan according to the histopathological diagnosis result can be implemented for a better after-surgery recovery. Moreover, data describing the outcomes for previous patients can be used to predict the recurrence rate or other potential adverse events. Therefore, the evaluation of the cancer regression grade is of vital importance for patients. Usually, there are two features that can be employed to evaluate

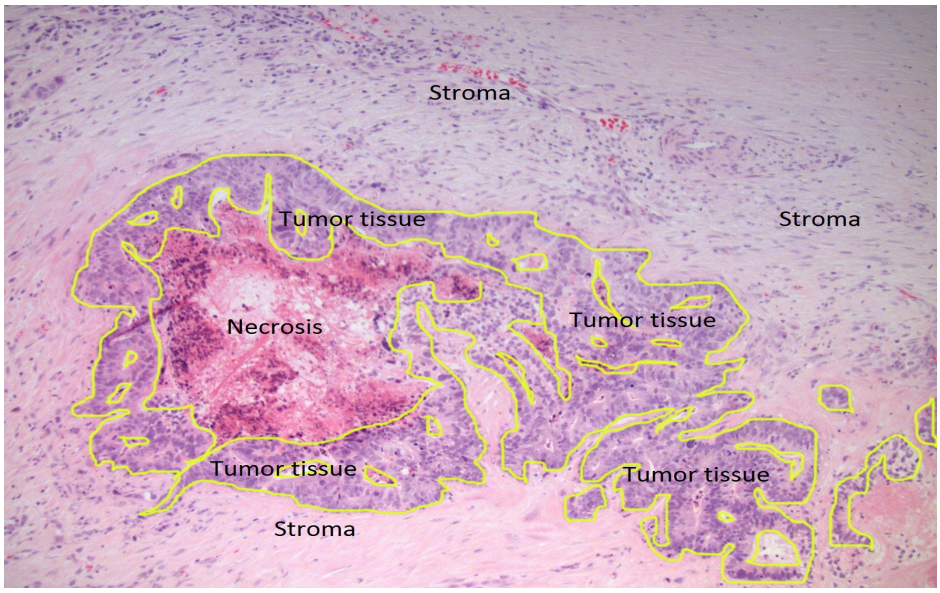


Fig. 1.2 Example of an annotated region for CRLM tumour regression assessment [3]

the regression. One is the percentage of residual tumour tissue in the whole observed area. And the other crucial visual clue is the distribution of normal cells and tumour cells around the invasive front [10]. The invasive front is the region where pro-tumour cells fight against the anti-tumour cells and can be presented through visual clues such as pushing, infiltrating or tumour budding. Fig. 1.2 demonstrates the images that are used for the tumour regression assessment through the proportion of the residual tumour cells on the whole region: the tissue that lies inside the yellow contour is tumour tissue; the tissue surrounded by the tumour tissue is necrosis killed by chemotherapy; and the other tissues with a lighter colour are normal

[3]This is cropped region of a WSI from CRLM patients and annotated by the experts

stroma. Fig. 1.3 illustrates different types of invasive fronts with the sketch and real patches from histopathology images. The different types of invasive fronts can also provide valuable information for clinical assessment. Currently, both kinds of histopathological evaluations are performed by expert pathologists through the visual assessment of the tumour slides. This is often inefficient and expensive due to the large number of slides that need to be reviewed and the limited availability of expert liver pathologists. Moreover, visual evaluations are inherently subject to inter- and intra-observer variability meaning that assessments can be unacceptably inconsistent and imprecise, and this can have a negative impact on the actual diagnosis and future treatment planning.

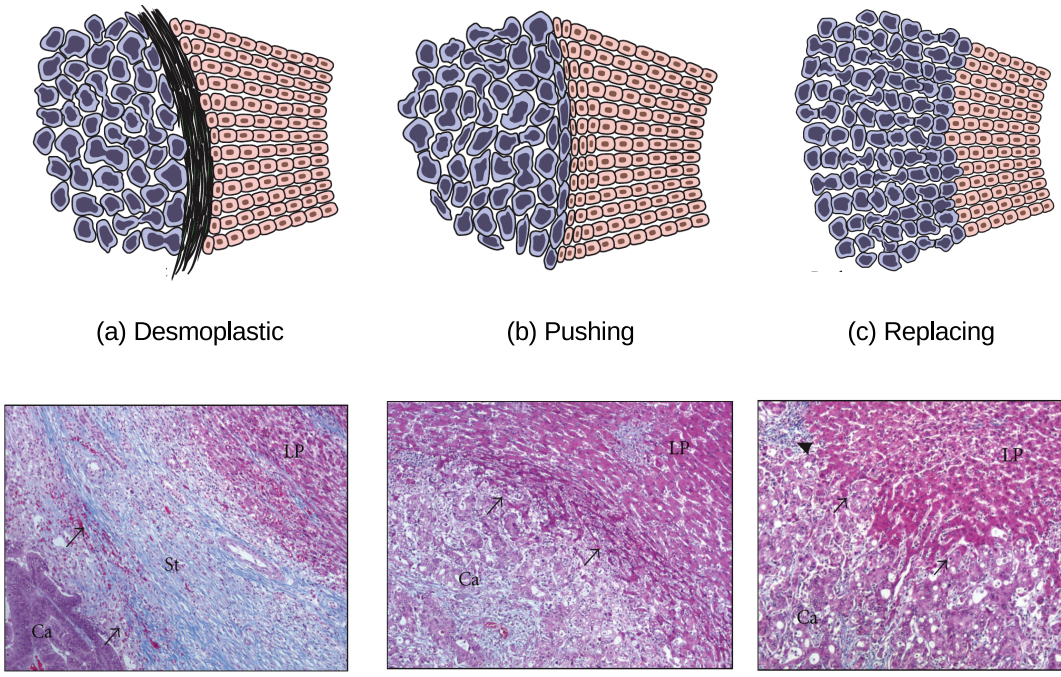


Fig. 1.3 Different invasive front patterns Images are adapted from Fig.1 in [11]. In the first row, blue irregular cells are the tumour, the orange ones are hepatocytes. In the corresponding H & E stained images, Ca denotes cancer cells while LP represents liver parenchyma.

1.1.3 Digital histopathology images

A sound understanding of the histopathology of whole slide images is the key to their successful analysis. Fig. 1.4 explains our understanding of the whole slide images. There are two different levels in this study, the physical level and semantic level. The physical level relates to the physical storage of the images. Due to the huge size of whole slide images (WSIs), WSIs are normally saved in a pyramid tiff format using tiles for the better performance of loading for visualisation. Subfigure (A) illustrates the pyramid structure of the file: the highest magnification level (in this study 40×) is located at the lowest level of the pyramid with the highest resolution; for each level up, the magnification level (whole image size) increases by a factor two. In other words, all the layers in the pyramid have the same content but different resolution.

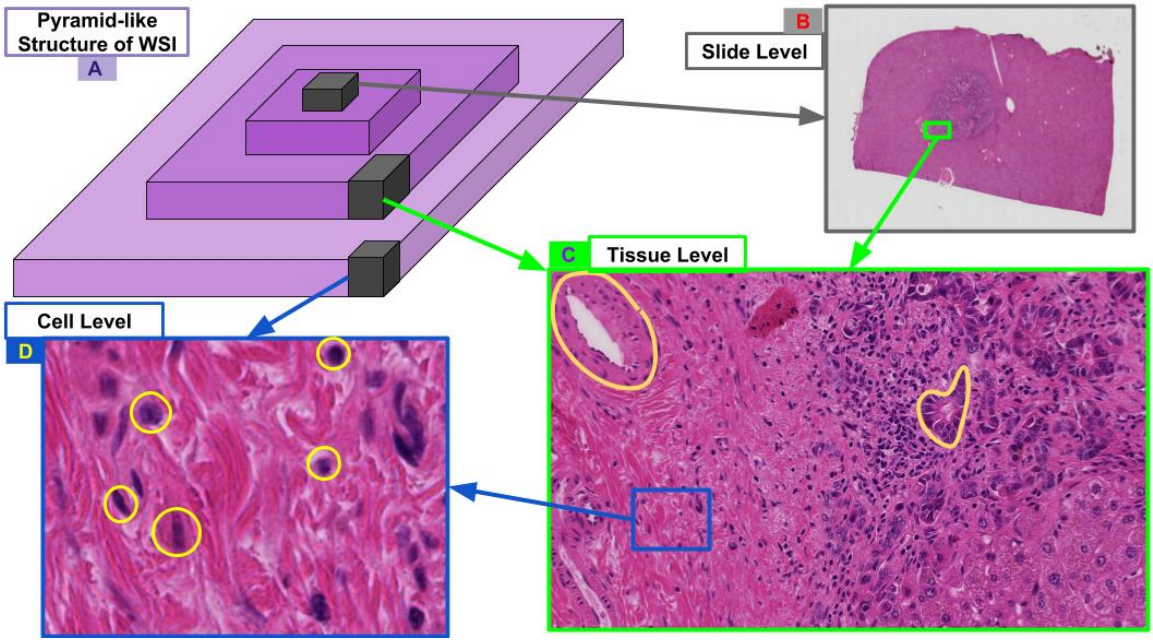


Fig. 1.4 Pyramid structure of WSIs and components observed on different levels

The subfigures (B-D) describe the semantic level of the whole slide. Although there are the associations shown in the figure, no fixed mapping exists between these two systems. At the semantic level, the contents in the scope are the main factors to consider. The slide level

(B) tends to consider all the contents in the slides. At the tissue level, there are some objects such as the glands and porta area that can be observed, except for different tissue components with their unique textures. The nuclei are the main objects that are distinguishable at the cell level, since the distribution of the nuclei is a matched reflection of the distribution of the cells.

1.2 Research objective

The demand for a computer-aided diagnosis (CAD) system for histopathologists to perform an objective, efficient, reproducible and quantitative analysis of a histopathology image has been put forward along with the emergence of digital histopathology images. Generally speaking, the study of an automated process to analyse histopathology images with computer vision is undertaken to assist biologists or doctors to accelerate the analysis process and save them from exhaustive and laborious manual inspection.

1.2.1 General tasks for histopathology image analysis

There are several general scenarios that the computer-aided techniques can help:

(a) Cells (nuclei) localisation and counting. Counting the cells or nuclei on a slide is a daily task for some biologists. For them, the exact number of objects has significant meaning. It seems to be an easy task for image processing, however, the challenges come when there are overlapping and blur in the image despite of huge morphology variance.

(b) Cells (nuclei) segmentation and classification. The classification of the cells (nuclei) is another meaningful task along with cell (nuclei) counting, especially when there are different cells mixed together. The cells (nuclei) segmentation place higher requirements on the computer vision task. To segment the cells (nuclei) out of the background

will help to gain a statistical understanding of them since there are masses of them in a region. The segmentation will also lead to a better classification performance since the foreground and background components are separated.

(c) Tissue level detection and classification. In the WSI, to detect and locate regions that have special structures like glands, invasive fronts, is another essential task. These kinds of regions are of greater interest to researchers and doctors, as both the number and the location of these special structures will provide valuable information about the whole slide.

(d) Tissue level segmentation for quantitative assessment. To analyse and segment different tissues in a whole slide image is a very challenging but important task especially for cancer-related diagnosis in clinical practice. Typically, the whole slide image for diagnosis purposes contains several different kinds of tissues such as normal tissue, cancerous tissue and dead tissue. In most cases, the percentage of tumour tissue in the whole slide is a critical factor. Yet, accurate calculation needs to be based on the accurate segmentation of the tissue.

(e) Virtual staining. The staining of the slides aims to highlight their different components. However, on the one hand, some of the histopathology images are grayscale images like those obtained with electron microscopy, to enrich the content and highlight different components, and in this case virtual staining could play a vital role. On the other hand, it is usually impossible to stain the slides with different dyes, whereas, with virtual staining, different components can be analysed by a computer and then stained.

The tasks above are only some common tasks that occur daily in a lab. With the development of computer vision and the improvement of a microscope, more and more application scenarios can be exploited.

1.2.2 Research objective of this study

In medical or clinical practice, one of the most important applications of histopathology image analysis is cancer grading. Histopathology image based cancer grading aims to detect and segment the normal and abnormal tissues or cells and then grade them based on medical guidelines. In this research, the overall objective is to develop CAD systems with intelligent algorithms to enhance the accuracy and efficiency of the automatic quantitative assessment of CRLM tumour regression and enable precise characterisation of the tumour invasion front. This objective requires development algorithms that have sufficient robustness and generalisability so that they can be applied to the segmentation and classification tasks of other histopathology images as well, especially for those with limited annotation data. More specific goals involved in this research can be broken down as follows:

1. **Classification:** develop computational models to classify the sub-types of tissue patches and nuclei based on the annotation information.

2. **Detection:** locate the regions of interest (ROI) in a larger patch or WSI, such as detecting the cancerous area in the WSI and detecting all the nuclei in an ROI.

3. **Segmentation:** segment the target tissue components (normally cancerous tissue) or nuclei out of the input image.

1.3 Challenges

Analysing histopathology images for a quantitative and reproducible result is a very challenging task for both pathologists and computers. The main drawbacks of the manual analysis conducted by pathologists' can be summarised as follows [2]:

1) Efficiency Pathologists need to manually zoom into different objectives and look through a WSI to complete their assessment. However, a WSI is huge, and the size of a compressed

WSI is normally several gigabytes, while on the $40\times$ objective, the size can go up to billions of pixels. Moreover, WSIs usually do not come alone, because diagnostic decisions need to be made on a series of slides. Hence, it can take hours for pathologists to analyse WSIs and make a final diagnosis for a single patient.

2) Reproducibility It is very difficult to define whether a tissue is tumorous or normal subjectively, especially for tissues in a transition state. Different pathologists may have their own rules to classify tissues, and the same pathologist may even make different decisions about a similar slide at different times. However, if the segmentation and classification are undertaken by algorithm, the results will not be affected by intra- or inter-observer subjectivity.

3) Quantifiability Quantitative analysis is usually conducted by the pathologist using computers for the task. For the pathologists, the evaluation of the percentage of the cancerous tissue on the whole slide is mostly based on visual estimation. While for the machine, the calculation of the index can go to the pixel level, even though it may not be necessary.

To overcome these difficulties and boost the efficiency of the analysis, pathologists have encouraged the development of a computer-aided diagnosis system for histopathology image analysis. Nonetheless, the task is very challenging for the computer as well, as unlike any other medical images or natural images, histopathology images have a unique visual characteristic. These difficulties can be summarised as follows:

1) Morphology The huge variance of morphologies at both the tissue and cell levels pose serious challenges for the automated analysis process. These variances not only come from different tissues that have totally different structures, but also from the same tissue because of the different sectioning angles. At the tissue level, due to a lack of shapes representing elements, there are few abstract representations for most tissue images. The visual appearance of the tissues at a lower magnification level is more like texture images, but the huge variance of morphology makes it different from most natural images. At the cell level with higher

magnification level, the outline of the nuclei can be easily observed. Despite the morphology of the nuclei at certain regions tending to be identical, there will be a blur in some areas and overlaps where there are high-density nuclei, which makes it difficult to detect and segment nuclei.

2) Big data with limited annotations The large amount of data itself poses a critical challenge, irrespective of the CAD process. On the other hand, annotations for these big data are considerably limited. The lack of adequate and high-quality annotations present significant obstacles for detailed analysis.

3) Human factor Different human factors during the preparation of the slides also increase the difficulties. The slide needs to be sliced and stained to highlight the biological structures such as the nucleus and stroma. The thickness of the slide, the die concentrate, as well as other factors will result in a huge difference between the slides. On the one hand, tissues can be very similar to each other, especially when they are at a transition stage. On the other hand, each pixel of the WSI needs to be assigned a label so that the area can be calculated for obtaining a quantitative result. These requirements add to the overall difficulty.

Despite all these difficulties and challenges, it is a promising research topic. More and more researchers all over the world are attracted to this field. Moreover, computer vision and artificial intelligence are finding their way to assist researchers and doctors to better observe and analyse histopathology images.

1.4 Outline of the thesis

Following this introductory chapter, **Chapter 2** presents a review of the methods for histopathology image analysis for both the cell level and tissue level. Notably, specific attention is paid to reviewing the application of deep learning methods on histopathology images. In addition, the new trend of using generative adversarial networks for weakly supervised learning and pixel level analysis is reviewed.

Chapter 3 addresses the task of the tissue level analysis with deep neural networks (DCNNs). Classification-based and segmentation-based methods are investigated to enhance the overall segmentation accuracy. During the processing, a number of feature maps and features are plotted to help understand how the networks make decisions. Variant approaches are conducted to visualise and evaluate the performance of different models on the WSIs.

Based on the rough analysis on the WSI, **Chapter 4** tends to further explore the information contained in histopathology images. Both supervised and unsupervised algorithms are applied to the histopathology for extracting cell level information. Based on the tissue level and cell level information acquired, a tissue region growing framework is proposed to enhance the segmentation results for the tissue border region.

In **Chapter 5**, another representation (staining) that highlights different tissue types of the CRLM WSI is introduced for the WSIs analysis. GANs are the main approaches facilitated to explore the inner connection between images with different staining. Aside from targeting at a better visualisation of the WSI, a weakly supervised framework is proposed to achieve the pixel level segmentation result.

In the last, **Chapter 6** presents the conclusions the findings and limitations of the work in this thesis, then proposals for future work are made.

Chapter 2

Literature Review

The study of computer-aided histopathology image processing dates back to the early 80s [12]. Although efforts have been made in this field, there are still very few commercial applications of automated histopathology image analysis that have reported until recently. With the rapid development of both software and hardware, especially with the help of deep learning, automated histology image analysis is eventually finding its way to clinical practice. Some early trials on specific cases have demonstrated its promising future, while more and more applications are on their way.

This chapter presents a review of the histopathology image analysis literature, with a particular focus on HE stained histopathology images. Firstly, based on the targeted contents of the WSI, this review is conducted from two different perspectives, the tissue level and cell level. Tissue level mostly summaries the methods that process the large patches as a whole. Whereas, the cell level examines the approaches that detect and segment the nuclei out of the complex background. Then the generative adversarial network (GAN) based algorithms, along with other weakly supervised methods are reviewed. In the final two sections, in addition to the introduction of the evaluation metric used for classification and segmentation, a few open-source tools and software are investigated for the purpose of integrating the algorithms into the platforms later.

2.1 Related work on tissue patch classification and segmentation

Due to the limitations of computation capability, the computer-aided processing of WSIs is normally based on tiles or patches, which only provide a partial scope of the whole slide. These image patches are called tissue level patches when they are cropped either with a large size (usually more than 400×400) on $40\times$ magnification level or relatively smaller size at a lower magnification level such as 200×200 at a $20 \times /10\times$ magnification level. The successful process of these patches at the tissue level is the groundwork for the WSIs analysis, and the review of the tissue level is conducted from two aspects, classification and segmentation based research.

2.1.1 Tissue level overview

Histopathology image patches are an import category of texture patches. Hence, most algorithms developed for extracting texture features can be directly used for histopathology images as well. However, there are some unique problems and challenges for histopathology image analysis, which are in addition to the challenges caused by the extremely large size of the WSIs. These challenges are described in detail as follows:

Problem 1 : what are the suitable patch sizes and magnification levels?

Finding a suitable combination of patch size and magnification level is a primary challenge in patch-based approaches. Fig. 2.1 illustrates the example of different magnification levels and different patch sizes. Apparently, with a lower magnification level, *e.g.* $10\times$, the details of the images are very blurry, and the nuclei can be observed as spots, but here, an overview of the bigger picture can be obtained. While on the $40\times$ level, the best image quality and details can be acquired, and texture of the nucleus and the cytoplasm can be observed. However, the

scope of the image is limited and the contextual information is ignored. The ignorance of

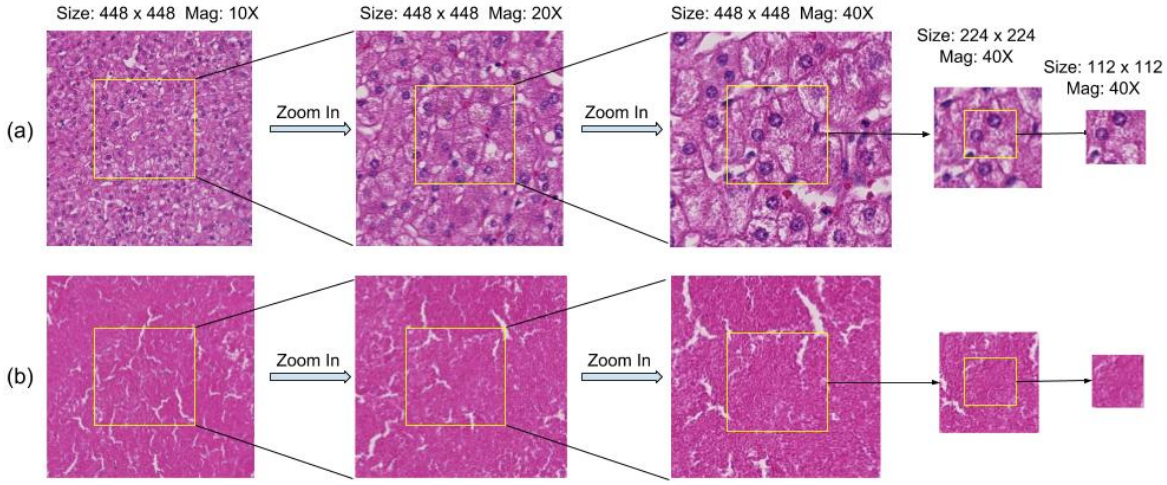


Fig. 2.1 Illustration of information contained in different magnification levels or patch size. (a): patches from hepatocyte with nuclei can be observed as individual objects in $40\times$ magnification level; (b) patches from necrosis demonstrate repetitive features across different magnification level.

context is even worse when the patch size becomes smaller. As shown in Fig. 2.1(a), for some kinds of tissue, *e.g.* hepatocyte, on $40\times$ magnification level, with a patch size $448 \times 448 px$, the patch can contain different nuclei which can be treated as individual objects with rich visual features such as the shape, size and filling content. However, without contextual information, it is difficult to judge which type of tissue this patch comes from, because such texture can exist in a variety of different tissue types. Moreover, for other types of tissues, as shown in Fig. 2.1(b), necrosis, the patches contain only textural features which are similar to noise. These textures are homogeneous irrespective of the patch sizes or magnification levels. In this case, the magnification level is not so important although the patch size could matter.

Problem 2: large scale dataset with repetitive features

One of the great challenges imposed on patch-based histopathology image processing is the vast scale of the data. For a large homogeneous region like 2000×2000 pixels, considering a patch size 224×224, with a step size 100, 400 training samples can be extracted without

involving any other augmentation procedure. For 10 WSIs with 10 annotations (the same size as described before) in each, a 40,000 training set will be obtained. However, it should be kept in mind that the actual sizes of the WSIs are normally greater than 1,000,000 $\times$ 1,000,000 pixels. To take into account larger areas in WSIs, the number of patches becomes enormous. Despite the advantage of abundant training data, there is a further obvious drawback: a large percentage of this data possess similar features, which may be even repetitive and redundant. The proportion of annotated data is another challenge that needs to be taken into consideration. Compared to the vast database collected by a different research institute, the portion of annotated data is certainly small. With limited knowledge provided to the model, methods to ensure the generalisability of trained models is always an issue.

Problem 3 : multi-class tissue classification and segmentation

Instead of performing a binary classification which mainly concentrates on detecting the cancerous area, the multi-class tissue classification and the segmentation task provide a more in-depth insight and offer a better understanding and assessment of histopathology images. The obscure border between different tissue types is one of the big obstacles for multi-class classification and segmentation methodologies. Some of the tissue regions are in the intermediate/ transformation state that no suitable class can be applied to, hence it is very difficult to define an clear border between different tissue. When it comes to segmentation, there is no explicit boundary that can be defined manually for training segmentation models. Another common issue is the data imbalance among different classes. The normal tissues account for a large portion of the samples, while the quantity of some other classes can be extremely rare, and this will cause a serious problem on the performance without proper handling.

Problem 4: rotate-invariant feature

The rotation-invariant feature is a critical feature for the histopathology image since the image actually comes from a 2D cross-section of a 3D object. Theoretically, there is no right direction for the histopathology images. This property can be useful on some occasions, such as, when augmenting the training dataset. Moreover, it brings potential problems. For example, if detecting a specific kind of structure in the whole slides is needed, the features from different angles will be individual which will enormously increase the workload.

2.1.2 Tissue level classification Method

The classification of the histopathology image patches is considered as the basic tasks for histopathology analysis. In this section, the trend of the classification methods applied on histopathology image patches is reviewed.

From handcraft features to deep features

In image processing based methods, feature extraction and classification are two independent tasks. Over the last few decades, most studies have concentrated on extracting the handcrafted features of the histopathology image patches. Texture features are absolutely one of the most important of these approaches. The Local Binary Pattern (LBP), Gabor filter, Haralick texture, and co-occurrence texture as well as many other texture features have been introduced or proposed to be applied to histopathology image analysis for various cancer types. The bag of features uses a feature dictionary learned from image patches and shows decent performance in many related kinds of research [13, 14]. Aside from the texture features, the intensity features, morphological features and topological features are very important for histopathology image analysis as well. Intensity features refer to utilising information from different colour spaces. Even though there is the staining variance issue, the colour of the whole slide images are constrained to limited subspace. Nonetheless, for most of the studies that have employed intensity features, maximising the intra-slide

variance and minimising the inter-slide variation are necessary procedures. In [15], the authors compare performances of 11 colour models in a classification task; they found that no model outperforms others in all cases. Whereas, the authors in [16] conclude that RGB colour space is better, while DiFranc et al. [17] state that adding CIELab information may be helpful. Morphological features focus more on the relative location and distribution, like the area, shape, perimeter, compactness, eccentricity & etc. [18] present an explicit shape descriptor and use it in histopathology image retrieval. Many other types of research [19, 20] have employed the gland area as an essential criterion in prostate cancer. Topological features which are also called graph-based features, can provide information about the structure of the sample. In [21], the nuclei were extracted and then used as the primary note for constructing the graph. The Voronoi Diagram, Delaunay triangulation, and minimum spanning tree are the most commonly used topological features for histopathology image analysis.

Deep learning algorithms have been the state-of-the-art methods for the natural image in different tasks such as classification and segmentation, and these have inspired researchers to apply them in the medical image field. Cruz-Roa et al. [22] present the first attempt of using a deep learning framework for the quantitative WSI analysis of invasive ductal carcinoma detection for breast cancer. By constructing a 3-layer CNN network, they obtained a 71.80% F-measurement and 84.23% balanced accuracy across 49 independent testing sets. Litjens et al [23] also employed the DCNN structure to detect metastasis in sentinel lymph nodes and to detect cancer in prostate slides, and demonstrated the powerful capability of the 'deep learning' technique in WSI analysis. Their team also hosted the CAMELYON16 (cancer metastasis detection in the lymph node) and CAMELYON17 contest, which has greatly attracted the interest of researchers all around the world and has boosted the usage of the deep learning algorithm with histological images. The winner of the CAMELYON16 [24] employed a much deeper network structure (GoogLeNet) and performed to the highest magnification objective ($40\times$), they obtained a 92.5% AUC (area under the receiver operating

curve) and 73.3% tumour localisation score. The author also compared the performance of other network structures including Alexnet and VGG16, but found that GoogLeNet yields the best training accuracy.

Nowadays, it is the trend to embrace deep features. Different studies have reported on different tasks through which the deep features can achieve much better results than the handcraft features. Obviously, it is not sufficient to only conduct the classification on the patch despite the excellent performance. Most clinical diagnosis is based on the WISs or a series of WSIs. Hence, focus on the slide level or case level is the next trend.

From tissue level to slide/case level Strictly speaking, there is no boundary between the tissue level and slide level/case level. Most slide/case level results are based on the output of the tissue level result, while the slide/case level result will try to output a semantic label of the whole slide instead of focusing on some of its parts. In the challenge of "CAMELYON", one goal of the tasks (CAMELYON16) is to distinguish whether the slide contains a cancerous area or not. Furthermore, in "CAMELON17", multiple slides from the same patients along with the patients' clinical data are provided. The objective is to predict a patient's pN-stage based on the detection and classification of metastases in multiple lymph node slides. During this challenge, many creative algorithms are developed to solve the two problems. Table 2.1 summaries some of the novel studies conducted based on this dataset.

From binary label to multi-type tissue classification

In terms of the environment of the tissue especially in the cancerous region, the components are very sophisticated. For example, in Fig. 2.2, from a patch cropped from the TCGA dataset; there are *Tumour, Fibrosis, Inflammation* and *Blood.* In a paper by Kather .*et.al*, a dataset contained eight different types of tissue patches from colorectal cancer (CRC) cases. In their work, they also provided a benchmark method by combining different handcraft features, which achieved 87.4% accuracy for eight classes. However, for this dataset, the patch size was too small, only $150px \times 150px$ at a $20\times$ magnification level, which will lead to

Table 2.1 Related research using CAMELYON dataset

Paper	Base Model	Novelty / Contribution
Wang [24]	GoogleNet, Alexnet, VGG16, FaceNet	Compare performances of different network structures
Liu [25]	Inception-V3	Compare performance from different magnification level
Yi [26]	Resnet	Introduce the conditional random field (CRF) for post processing
Lin [27]	VGG-based fully connected network	Fast processing ability with large patch size
Courtiol [28]	ResNet-50	CHOWDER as a weakly-supervised architecture
Agarwalla [29]	Alexnet	A context-aggregation network with 2D-LSTMs
Kohl [30]	DenseNet	Transfer learning for better performance

insufficient information for classification (especially for contextual details). Also, the dataset mostly comes from one lab meaning that it would not contain enough staining variance to build a robust and universal model.

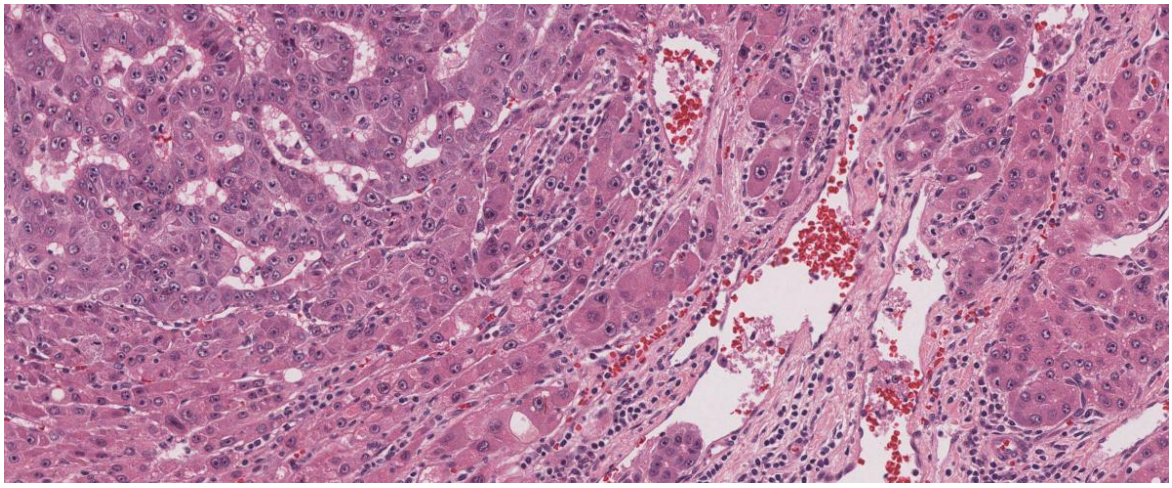


Fig. 2.2 Example of tissue patch with different components from TCGA dataset

From small patches to large patches

During the first task of the ICAR2018 Challenge- BACH, another dataset with four different

types of tissues labelled for breast cancer was introduced. Each patch was annotated with a patch-wise label: normal, benign, situ carcinoma or invasive carcinoma. Yet, the patch size for each patch can go up to $2048px \times 1536px$ for each training image. More detailed information can be found in the paper by Araújo et al. [31]. However, one drawback of the contest is that the total size of the provided dataset is not large enough compared to the other contests such as CAMELYON. A critical feature of the dataset is that it comes with a staining variance which adds to the overall difficulties. There usually are two ways to deal with this issue. Some researchers claim that image pre-processes with the stain normalisation method are not necessary with proper colour augmentation, while others insist on applying a colour normalisation method before training. By any means, the purpose is to deal with the staining variance among different patches from different slides. The best option might be to keep the data variances during the training to make sure the networks learn the colour-invariant features and apply a normalisation method before testing to ensure robust performance.

Table 2.2 A summary of the methods employed for BACH contest

Paper	Patch Size (px)	Base Network	Normalisation	Pre-trained Model
Araújo et al. [31]	512×512	8-layer DCNN	Yes	No
Nazeri et al. [32]	512×512	FCN	No	No
Vang et al. [33]	512×512	Inception-v3	Yes	Yes
Vesal et al. [34]	512×512	Inception-V3,Resnet	Yes	Yes
Xie et al. [35]	512×512	DEN+SCU	No	No
Awan et al. [36]	512×512	Resnet	Yes	No
Rakhlin et al. [37]	400×400 , 650×650	ResNet-50, InceptionV3 and VGG-16	No	Yes
Golatkar et al. [38]	299×299	Inception-v3	No	Yes
Cao et al. [39]	224×224	A multi-view RFSVM method	No	Yes
Safwan et al. [40]	224×224	Resnet	Yes	No

Regarding network structures, generic deeper networks such as the inception net and Resnet are more popular compared to networks specially designed for texture images.

Nonetheless, transfer learning is also preferable to training from scratch. Besides, because only one label is provided for the large patch in the training dataset, a more substantial patch size such as ($512px \times 512px$) is used for the network input. To make a single prediction based on the extensive information for a $2048px \times 1536px$ image patch, more algorithms focus on integrating the contextual information or for fusing the results from the different levels or different networks. Cao et al. [39] proposed a multi-view RFSVM method to fuse the results from different models. In the work of Rakhlin [37], networks with different input patch size are fused to gain better performance, while Awan et al. [36] have attempted to integrate contextual information. Other methods such as the "Dual Path Network" proposed by Vang et al., "Two stage network" by Nazeri et al., and "Reversed Active Learning" by Xie et al. have also been proposed to refine the results from small patches.

2.1.3 Tissue level segmentation methods

At the tissue level, there exist two kinds of segmentation: for those with a particular type of structure such as the gland, the segmentation can be treated as object-oriented segmentation. For the others that contain only textural features, the segmentation can be considered as textural segmentation.

The GlaS (Gland Segmentation) contest [41] hosted by Warwick University is focused on the first kind of segmentation which is another critical contest that significantly boosts the application of the deep learning algorithm on histopathology images. The objective of this contest is to achieve semantic segmentation over the glands area for colon cancer histopathology images. Dozens of deep learning methods have been proposed for this task [42, 43]. Some researchers have tried to incorporate the features from different perspectives for a better result. [44] propose a deep contour-aware network based on the fully convolutional network, which can incorporate multi-level contextual features for an end-to-end result. Xu et al. [45] present a deep multichannel side supervision network that can harness the features of both

edges and the blocks, and this leads to better and more accurate border info. BenTaieb et al. [46] propose a topology-aware fully convolutional networks to segment the gland area. In their study, they propose a new loss to include geometric and topological constraints into the deep network. They compare their results with several segmentation methods for natural image segmentation such as FCN-32s [47], Deeplab [48] and CRF-CNN [49]. MIMONet proposed by Raza et al. attempts to use multiple-input and multiple-output at the same time and achieved state-of-art results [50]. Other studies such as [42], [51] & *etc.* seek to merge information from different perspectives as well. At the same time, some researches have tried to employ other methods to overcome the limited amount of ground truth provided. For example, in the studies by Yang et al. [52] and Zhang et al. [53], they attempted to utilise unannotated data to augment the training. Nonetheless, Singh et al. [54] present the whole workflow for gland segmentation in the prostate including the annotation procedures.

Apart from the studies based on the GlaS dataset, semantic segmentation of other cancer types targeting at texture-based segmentation is more challenging. In the paper by Wang et al. [24], they introduce the fully convolutional network (FCN) for the semantic segmentation of muscle and messy regions. Aside from this, the recurrent neural network (RNN) was firstly introduced for histopathology image segmentation by [55]. The authors in this paper proposed a 2D spatial clock-wise RNN model for muscle perimysium segmentation. The comparison with U-net demonstrates the network to be a promising structure for generic histopathology image segmentation. Instead of using a sequential network, Gadermayr et al. employed a cascaded strategy to organise a different U-net based sub-network. Another end-to-end network described by Mehta et al. [56], uses the ground truth label on the pixel level, while the input patches contain a variety of types of breast cancer. To tackle the problem, the authors constructed an encoder-decoder network structure which can also fuse output from the multi-scale input to obtain an end-to-end result.

2.1.4 Publicly available datasets on tissue/slide level

A high-quality annotated dataset is a critical prerequisite for training a good deep learning model. However, the annotations for these histopathology images are very difficult to obtain. Unlike the annotations for natural images which can be done by a normal person with little training, the annotation for medical images needs to be conducted by experts with corresponding medical knowledge. Table 2.3 lists a few publicly available data sources on the tissue level.

Table 2.3 Public dataset for tissue-level histopathology image analysis

Dataset	Description	Size	Annotation
TCGA [1]	Multiple different cases	more than 1000 WSIs	No annotation
TCIA [2]	Multiple different cases	more than 1000 WSIs	No annotation
GlaS [3]	Gland area segmentation for colon tissue	50 ROIs with size 600×400 for training, 46 testing	Gland areas
CAMELYON-16 [4]	Breast cancer lymph node matestasis	270 WSIs for training, 130 WSIs for testing	Tumor area
TUPAC16 [5]	Breast cancer cases	500 WSIs for training	Proliferation scores based on mitosis counting
CAMELYON-17 [6]	Breast cancer lymph node matestasis	500 WSIs for training, 500 WSIs for testing	Lesion-level and patient-level annotations
BACH [7]	Breast cancer cases	400 ROIs with size 2048×1536 for training	Patch-level label with 4 different classes

[1] Webpage: https://cancergenome.nih.gov/
[2] Webpage: http://www.cancerimagingarchive.net/
[3] Webpage: http://www2.warwick.ac.uk/fac/sci/dcs/research/tia/glascontest/
[4] Webpage: https://camelyon16.grand-challenge.org
[5] Webpage: https://camelyon17.grand-challenge.org/
[6] This challenge is an extension task of mitosis detection. Webpage: http://tupac.tue-image.nl/node/3
[7] Webpage: https://iciar2018-challenge.grand-challenge.org/Dataset/

2.2 Related work on cell level tasks

In HE stained histopathology images, the morphology of the nuclei can be clearly observed with a more than 40× magnification objective. Even with a 20× magnification level, the outline of the nuclei can be easily distinguished except under a few complicated circumstances. Hence, at the cell level, in most situations, instead of detecting and segmenting the whole cell, the nuclei are the main targets. The properties of the nuclei, such as its shape, size and distribution are important properties for differentiating cells, and is important for biomedical analysis. This is because cells are the basic functional units of the living creature, and the nuclei are the core part of the cells.

Nuclei detection, classification and segmentation were well-studied fields before the trend of deep learning began. Dozens of methods have been proposed to tackle these tasks over the past few decades. Irshad et al. [57] presented a review on the methods for nuclei detection, segmentation and classification in 2014, in which the author reviewed most of the start-of-art image processing methods for cell level tasks including threshold, morphology, region growing, watershed, active contour, k-means clustering, probabilistic models and graphic cuts. The authors also provide a table which describe the datasets, objectives and metrics used in previous studies. In recent years, inspired by the great success of deep learning algorithms on natural image processing, more and deeper learning algorithms have been introduced or designed for nuclei detection, classification and segmentation. A few recent works are described in the survey by Litjens et al. [58]. The deep convolutional neural network, stacked auto-encoder and the recurrent neural network are the most adopted methods employed for cell level tasks. In this section, the work on the cell level are summarised from four aspects as detailed in the following.

2.2.1 Image pre-processing: image deconvolution and normalisation

With more and deeper learning models applied to histopathology images, image normalisation is no longer a necessary procedure to obtain a robust result. However, due to the huge variance of the images caused by different concentrations of the dye and the different illuminant in different labs, typically, image normalisation is normally the first step before other image processing algorithms can be applied. Although, it is also essential to enhance the visual experience of the captured histopathology images on the screen for the histologists. In addition, colour deconvolution is the usual approach for colour normalisation, which separates the different channels in the original image in terms of the colour of the stains. With the separated H channel image, the nuclei can be easily located and segmented with a threshold. Hence, a brief introduction of colour normalisation methods is provided here.

Typically, there are only limited colour components presented in one stained histopathology image. For example, there are only two main colours in the HE stained images, blue and magenta/pink. The purpose of colour normalisation is to align the main colour components in different images to realise a universal appearance. Li and Plataniotis [59] have summarised the previous work on image normalisation into three categories: histogram matching, colour transfer and spectral matching. They also propose a novel framework which attempts to address the three challenges (illumination and stain variation, preservation of histological information, and the requirement of a reference image) in one framework. Janowczyk et al. conclude that previous work can be grouped into a stain specific algorithm and clustering type algorithm [60]. Then, based on Basavanhally and Madabhushi's work [61], the authors propose a StaNoSA model which utilises a sparse auto-encoder to extract the features from both the reference image and the target image before applying the alignment.

2.2.2 Nuclei detection

The basic goal of nuclei detection is to initialise the seeds of possible locations of nuclei which is critical for subsequent biomedical analysis. However, more advance goals require accurate localisation of the nuclei, such as in a bounding box way. Conventional nuclei detection methods prefer to use the statistical or geographical properties of an image to generate the seeds. After the image normalisation, the most straightforward method to locate the seed position for the nuclei is the threshold. More advanced method such as the H-maxima transform [62, 63], Hough transform [64], and radial symmetry transform (RST) [65, 66] detect seed points with local peak or shape information & et al. The voting method proposed by Qi et al. [67] utilises the shifted Gaussian Kernel for single-path voting and demonstrates a better performance than the iterative voting method proposed by Parvin *et.al* [68]. Most image processing based methods can achieve decent results with normal tissues where the nuclei distribute uniformly. When it comes to a complicated situation like tumour tissues that have crowded and overlapped nuclei, they may not work as well. Moreover, they are not robust enough since the performances are subject to the appropriate parameters being selected. With the wide application of the deep learning method, the situation is constantly evolving, and more and more robust and accurate algorithms are proposed for cell level tasks.

The general deep learning framework normally employs CNN to generate feature maps or to build a regression network directly. The work by Xie et al. [69] has proposed a novel fully convolutional regression network structure for an end-to-end prediction of the central point of cells. Their results indicate that the proposed method has a good performance in terms of overlapping and clumping cells. Instead of using the central location of the cells, another regression network represented in [70] employs the bounding box for cell(nucleus) detection which increases the overall difficulty by introducing another dimension to regress. But the outcome shows an excellent result, especially for the overlapped cells. To adapt the CNN to suit the nuclei detection task, the authors in [71] propose a spatially constrained network

which merges the spatial info into the network. Usually, using deep learning methods takes a long time to process a large image. While Xu and Huang [72] propose a distributed network with the pre-fetch operation and compressed network parameters to accelerated the process of detecting the nuclei on the whole slide image. As their title states, they are able to detect 10,000 cells in one second. Inspired by the voting method mentioned before, another voting method which combines the CNN and the Hough voting method was proposed by Xie et al. The advantage of this method requires relatively low efforts of annotation. Other methods that utilize unsupervised networks or semi-supervised networks have also been proposed. Xu et al. [73] tried to extract the features of the nuclei with an unsupervised network - the stacked sparse auto-encoder, and then used the extracted features to classify the foreground and background on breast cancer histopathology images with 3,500 manually annotated ground truth.

For most image processing based methods, they do not require a large training dataset because they are mainly based on prior knowledge. While, for most deep learning based methods, a large well-annotated dataset is necessary. Hence, even though the deep learning methods for this task demonstrate excellent performances, methods to efficiently take advantage of prior knowledge may be the next trend in the development of this task.

2.2.3 Nuclei segmentation

In previous research, the pure segmentation method only considered the edge information and morphology features. Hence, the performance for overlapped nuclei or cells is not satisfactory. A more reasonable and high accurate segmentation, semantic segmentation and instance segmentation need to be considered. Watershed [74, 62], active contour [64, 75] and graphical cut [76] were previously the fundamental methods for nuclei segmentation. Many advanced methods and frameworks [77–80] have been developed based on these methods.

These methods are proposed to overcome the disadvantages of the basic methods and tackle the difficulties for the overlapped nuclei.

To adopt a deep learning method for the nuclei segmentation task, the normal DCNN model is not suitable for the task. Ronneberger et al. [81] have proposed a specially designed U-net structure for biomedical image segmentation which has proven to be very successful for many related tasks. While Akram et al. [70] fused the extracted bounding box information with another network for the segmentation task. The authors in [82], attempted to integrate the information from multi-scale, and also incorporated high-level prior information to reconstruct the border of the overlapped cell(nuclei). The work in [83] utilized the probability map generated with DCCN to initialize the shape, and then the final segmentation would be completed by a deformable model and a sparse shape model. Their team also proposed a fast-scan CNN for region segmentation [84].

2.2.4 Nuclei classification

Nuclei detection and classification are two closely related research topics. For example, the detection problem can be converted into a binary classification problem with the label background or foreground. Furthermore, for some deep learning frameworks, the detection and the classification tasks are deeply integrated into one network.

However, before the deep learning dominates the computer vision related task, the nuclei detection, feature extraction and the classification are independent tasks. The shape, size, and smoothness of the cell (nuclei) are important visual features for the classification. Other than that, spatial information such as the density, overlap ratio & *etc.* are also important clues. These features are intensively used in [85, 64]. In 2012, a challenge to detect mitosis in breast cancer histological images was proposed [86]. The winner using a CNN model beat others who used traditional methods and achieved a big advantage. Afterwards, many methods using the deep neural network were proposed [87, 88].

Table 2.4 Public datasets for cell-level histopathology image analysis

Dataset	Descriptions	Size of the dataset	Annotation
MITOS-12 [8]	Mitosis detection for breast cancer evaluation	50 high power field image on 40× magnification level with patch size more than $2000px \times 2000px$	Center of each Mitos
AMIDA13 [9]	Mitosis detection for breast cancer evaluation	21 patches with size $2000px \times 2000px$ on 40× magnification level	Centre of each Mitos
MITOS-14 [10]	Mitosis detection for breast cancer evaluation	284 frames at 20 × magnification and 1,136 frames at 40 × magnification	Center of each Mitos
CASE-Nuclei Detection [11]	Breast cancer image for nuclei detection	Training: 14421 nuclei patches and 28032 non-nuclei patches with the size $68px \times 68px$ Testing: 516 ROI with size $400pxpx$ are annotated	Central point of the nuclei
MoNuSeg Challenge [12]	Multi-orgran nuclei instance segmentation	32 images with size $400px \times 400px$ for training, 14 images for testing	Mask of each nuclei
Kaggle Nuclei Segmentation Challenge [13]	Patches from various microscopy images	670 images with size varies from $256px \times 256px$ to $600px \times 1000px$ for training; around 1000 patches for testing	Instance-level annotations

Except for detecting and classifying specific types of cells, classifying different sub-types of cells in one image is also a popular research topic. Yao et al. [89] leveraged the features extracted with DCNN to differentiate biomarkers in lung cancer and used them for survival prediction. Based on the spatial constrained network, [90] proposed a novel Neighbouring Ensemble Predictor (NEP) to classify the four types of nuclei in colon cancer. To solve the

[8] webpage: http://ludo17.free.fr/mitos_2012/download.html
[9] webpage:https://ismi-amida13.grand-challenge.org/
[10] webpage: https://mitos-atypia-14.grand-challenge.org/home/
[11] webpage:http://gleason.case.edu/webdata/TMI2015
[12] webpage: https://monuseg.grand-challenge.org/
[13] webpage: https://www.kaggle.com/c/data-science-bowl-2018/

insufficient annotations problem, the authors in [91] proposed an AggNet with a new concept - learning from crowds. Jacobs et al. [92] compared the performances of transfer learning and the ladder network for detecting and classifying different nuclei with limited data.

2.2.5 Publicly available datasets for cell level

Over the past decades, although a number of datasets have been published, few of them have high-quality images and detailed annotations. For example, some of the annotations just mark out the rough central location of the nuclei with a point. Another common problem is the number of collected images in the datasets. Due to the high density of the cells, it is very time-consuming to collect a large dataset, and is probably the reason why high-quality cell level datasets are so rare. In this section, a few publicly available datasets with good image quality and annotations are considered. They are listed in Table 2.4 with their website.

2.3 Towards a pixel level analysis with weakly-supervised methods

Generative adversarial networks were first introduced in 2014 by Goodfellow et al. [93] to model the distribution of input data (image). As an important branch of deep learning, GANs have gained more and more attention from researchers around the world since their presentation. Normally, the generative model is trained to learn the underlying distribution of the data instead of memorising the entire data. More specifically, the objective of GAN is to learn those specific semantic and structural properties and then create new reasonable samples from the trained model. Hence, a significant amount of investigations have explored the potential of GANs for the generation of realistic images [94–100]. On the other hand, with some additional input constraints, the GANs can be applied in more promising tasks like image super-resolution, style transfer, and semi-supervised learning, amongst many

others [101–108]. In the review by Xin et al. [109], recent medical related research using GANs was summarised.

2.3.1 GAN-based networks for dataset augmentation

One of the main challenges for histopathology image analysis is the insufficient annotations available for training, since the annotations for the medical image have to be performed by clinical experts, which is costly and time-consuming. In this case, generative methods can play an important role in augmenting a training dataset to achieve a decent final result. A few initial efforts were made to explore the possibilities of applying this on biomedical images. For example, Osokin et al. used GANs for synthesising cells imaged by fluorescence microscopy [100], their results revealed that GANs are able to "simulate the localisation of multiple proteins throughout the cell cycle"; Costa et al. [94] and Guibas et al. [110] synthesise retinal images by applying adversarial training.

Most current adversarial training based image synthesis systems try to generate realistic images from a random noise vector, or by refining low-quality images to create more realistic high-quality versions. However, modelling the tissue patterns of different categories in histopathology images is a particularly complex task for generative models since they demonstrate various complicated morphological features. There are several challenges to synthesising histopathology images. The current image synthesis approaches often require a physics-based 3D construction and rendering model [99], but the physics at the cellular level is mainly unknown, making physics-based modelling unfeasible. The CytoGAN model proposed by Goldsborough [111] attempts to synthesise new cells by combining the different morphology of different cells. In the work of Mahmood et al. [112], they tried to generate realistic histopathology images from generated masks which will help to create ground truth for nuclei segmentation at the same time.

Histopathology images are rich in terms of their structure and texture characteristics when observed at different levels of magnification. It is hard to synthesise a whole slide image with a large variety of morphology features. Moreover, care must be taken to avoid bias in synthesised images. Most synthesised histopathology images focus on synthesising a gland area or nuclei related regions due to their multi-scale, random distribution and rotate-invariant features. More applications of adversarial augmentation data for improving model performances on histopathology images were introduced in [113–117], the results demonstrated the feasibility of using GANs for data augmentation.

2.3.2 GAN-based networks for image enhancement

Image enhancement is another important research topic through which GAN related networks can help. Methods to improve the visualisation of histopathology images is at the core of the challenge due to their crucial impact on the efficiency of pathology examination and diagnostic accuracy. Following novel GANs applications for natural image processing, virtual staining and super-resolution are promising and related tasks for histopathology image analysis.

Research on virtual staining has been conducted for many years and applied to many different types of images. The recent advances in GAN based approaches with their superior abilities have opened new roads in this direction. Early virtual staining research mainly focuses on the low visual level, for instance, image normalisation can be treated as a special type of virtual staining. In the research by [118], a GAN-based network is proposed for normalising the input images. While other research will be more straightforwardly related to virtual staining in terms of the concepts. Rana et al. [119] present a GAN-based network to stain and de-stain the prostate biopsy. Histopathology image super-resolution tasks can be treated as low visual level tasks as well. Despite extensive research on natural image processing tasks [120, 101], few studies have explored the potential to apply them to

histopathology images. No matter whether the image is enlarged for a more detailed view or the image is compressed for a more compact size, the GAN will play a significant role.

More advanced (high level) virtual staining will involve a transformation in terms of the semantic understanding of the region which may contain different tissue components. Recent work has demonstrated that with the help of deep learning, different tissue components can be separated semantically. This means that the staining of different tissue components can be separately and re-colourized regardless of the original staining technique. Early trials from Bayramoglu et al. [96] utilised a Conditional Generative Adversarial Networks (cGAN) to virtually stain unstained specimens and Rivenson et al. [121] employed the GAN model to virtually stain the fluorescence images to HE images. No GAN-based semantic virtual staining, which is considered a deeper process for managing raw images and is related to image segmentation and classification tasks have been reported yet. Although a few attempts have been reported using deep neural networks. Trahearn et al. [122] propose a hyper stain inspector for image alignment and cancer detection. Another innovative way "virtual staining" method has been proposed in [123]. The authors represent different tissue components in an image with colourful plates of different sizes, thereby facilitating diagnosis.

2.4 Evaluation

Proper evaluation is of significant meaning for both algorithm design and clinical decision-making. The evaluations in this thesis are mainly carried out from two aspects: classification and segmentation.

2.4.1 Evaluation metrics for classification

For classification-oriented tasks, the main metrics concerned are **Accuracy**, **Precision**, **Sensitivity**, and **F-Measures (FM)**. With true positive denoted as TP, false positive as FP, true negative as TN and false negative as FN, the above metrics can be described as following:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{2.1}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{2.2}$$

$$Precision = \frac{TP}{TP+FP} \tag{2.3}$$

$$FM = (1+\beta^2) \times \frac{Precision \times Sentivity}{\beta^2 \times Precision + Sensitivity} \tag{2.4}$$

where β is a coefficient for FM, typically, 1 is the most commonly applied value for β. In this case, the F-measure will be called the F1-measure or F1-score. For algorithm performance evaluation, accuracy and the F1-score are the most commonly applied metrics. However, in clinical practice, precision and sensitivity are more important indexes for doctors.

2.4.2 Evaluation metrics for segmentation

For segmentation-oriented evaluation, **Pixel Accuracy (PA)**, **Mean Intersection of Union (MIoU)** and **Dice indices** are employed in this study. The PA is calculated based on the predicted value at the pixel level. The pixel accuracy has the same definition as the accuracy in classification-oriented evaluation shown in Eq. 2.1. However, the definition of TP, TN, FN and FP are based on each pixel.

The MIoU is the mean of a series of Jaccard indices, since Jaccard indices are called the intersection over union (IoU). Given the segmented result S and ground truth G, the definition of Jaccard and Dice indices can be written as :

$$Jaccard(G,S) = \frac{|G \cap S|}{|G \cup S|} \tag{2.5}$$

$$Dice(G,S) = 2 \times \frac{|G \cap S|}{|G| + |S|} \tag{2.6}$$

Hence, the MIoU can be defined as:

$$MIoU = \frac{1}{c} \sum_{i=1}^{c} (\frac{1}{n} \sum_{j=1}^{n} IoU_{(i,j)}) \tag{2.7}$$

where c is the number of classes and n is the number of samples in each class.

Another perspective from which to evaluate the segmentation result is to measure the boundary similarity. For example, in the GlaS contest [41], they introduced an objective-level Hausdorff distance to evaluate the boundary accuracy. However, for the histopathology image segmentation at the tissue level, it is not as important to evaluate the boundary accuracy since it is very difficult to get a precise annotation in the border region. Following the semantic segmentation survey for natural images using a deep learning method as described in [124], no boundary-based metrics have been reviewed.

2.4.3 Other metrics

The receiver operating characteristic (ROC) curve is a popular metric which provides a graphical explanation for a binary classifier, and can be used to evaluate the performance of both the classification and segmentation tasks. The ROC curve plots the true positive rate along the false positive rate for different cut-off points. The points on the ROC curve denote sensitivity pairs for a particular decision threshold. The area under the ROC curve (AUC) is another important measure which indicates how well the model can distinguish between two

different groups. Mean average precision (mAP) is an important metric for object detection, especially for the nuclei detection task. The average precision (AP) is a calculation based on the area under the curve (AUC) for the Precision-Recall curve. During the calculation of the precision and recall, the detected mask will be considered true if the Jaccard index is greater than a threshold (0.5). While, the mAP is the mean AP value through all the samples in the evaluation set.

2.5 Open-source platforms for histopathology images management and analysis

To manage the increasingly large amount of histopathology image data, a few comprehensive and powerful platforms have been developed to achieve efficient data management and cooperative annotation.

2.5.1 Traditional cell level platforms

Before the wide application of the digital scanner of histopathology image and the explosive increase of the WSIs, researchers mostly focused on the processing of the cell level images. A few tools developed for cell or nuclei processing (cytology) were reviewed in [125, 126]. Among them ImageJ [127] and Cellprofiler [128] are the most powerful. For whole slide image management, annotation and analysis, many platforms have been built in recent years. Cytomine [129, 130] and Qupath [131] are two such examples.

ImageJ/Fiji[14] ImageJ is a Java-based medical image processing software. Its biggest advantage lies in its rich resources of plugins which can be easily utilized by other software. Furthermore, the large user community provides great support.

[14] https://github.com/imagej/imagej

Cellprofiler[15] Cellprofiler is developed with MatLab. Initially, it was designed for biologists to quantitatively process the cell in the images. CellProfiler Analyst [132, 133] is an upgraded version which integrates several machine learning algorithms to enable the interactive analysis of data.

Open Microscopy Environment (OME)[134] For the whole slide image management and analysis, OME has existed for a long time. It is a powerful tool, and supports different formats which include the most commonly used formats. Moreover, they have different products for different application scenarios, which include both commercial and non-commercial products. In a client-server manner, all the data are stored in the server to ensure the safety of the data and save the local storage space.

2.5.2 Basic libraries or tools

In contrast to the commonly used image formats, such as *.jpeg*, *.png*, the digital pathology image needs to handle extra-large images in terms of pixel size. Normal software cannot handle these kinds of images directly, and special tools are needed to decode the image.

Openslide [135] The WSIs used in different hospitals/countries are not in a universal format, and for commercial reasons, different scanner manufacturer companies have their own protocols to encode and decode their images. With the rapid growth of histopathology image usage in both clinical practice and scientific research, Openslide, as an open source application interface with integrated multiple format support for WSIs, has been developed to deal with this situation. Openslide is the base for many other platforms such as ASAP, Cytomine and Qupath.

Libvips[16] is another powerful library for histopathology whole slide image processing, which can be treated as an updated version of Openslide. Libvips is built for efficient large scale image processing with low memory cost. However, it is not only used for histopathology

[15] https://github.com/CellProfiler/CellProfiler
[16] https://jcupitt.github.io/libvips/

images but can also be applied to any large scale images. Many advanced operations are also supported such as rotating, flipping, and format converting.

Slidealtas[17] is more of a web-based visualisation tools/script than a basic library for histopathology image processing. It is mainly used in web-based visualisation tools for loading image tiles from the server according to the users' requests. In contrast to other libraries such as Openseadragon, Openlayers, Leaflet and Geojs which have been developed mainly for geographical related tasks, slidealtas is designed especially for histopathology image visualization.

2.5.3 PC-based software

There are many PC-based software packages available on the market, mostly from scanner providers. However, one of the main drawbacks of these software are that they only support their own formats. Moreover, this is the reason why the open-source platforms is more popular.

Automated Slide Analysis Platform(ASAP)[18] ASAP is a light-weight open source application for WSIs visualising and annotation which is based on Openslide. The Qt-based viewer inside supports fast and smooth visualisation, and supports saving large images to generic multi-resolution tiled TIFF files.

Qupath [131] QuPath is specially designed software for whole slide image analysis and digital pathology. Since it is a Java-based application as well, it provides good support for different kinds of extensions, for example, the ImageJ is included in the software.

[17] https://slide-atlas.org/

[18] https://github.com/computationalpathologygroup/ASAP

2.5.4 Web-based platforms

With the growing amount of slides scanned in a hospital every day, the big data problem is a serious concern. This also indicates that a web-based system which stores all the slides in the cloud is more suitable for such tasks.

HistomicsTK [19] HistomicsTK is both a pure python toolkit and web-based tool for histopathology image processing and analysis. With respect to the web platform component, it serves as a plugin for Girder which is a very powerful data management platform. Girder supports different plugins including Openslide. Hence, it can be used as a histopathology image organization and dissemination platform with user management.

Cytomine [136] Cytomine aims for a collaborative web-based platform, where the web interface will ensure the same user experience across different operating systems. The whole system is made of four main components. The core is in charge of the overall logic of the system and responds to the requests from users. The image management system (IMS) manages all the image files. The WebUI module is the interface provided to the users, where users can upload, visualise, and annotate the image. The last main module is the Data Mining module which includes different machine learning algorithms or other retrieval algorithms. Another main feature of the platform is that these modules are independent of each other, and can easily be deployed to different servers to increase the overall performance.

[19] https://github.com/DigitalSlideArchive/HistomicsTK

Chapter 3

Deep Neural Networks for Histopathology Whole Slide Image Classification and Segmentation

In the computational analysis of histopathology WSIs, extremely high-resolution images are often decomposed and analysed in the unit of patches that have a more practical size. This kind of patch-based approach is particularly important when WSIs are to be processed with DCNNs that can normally only take a fixed size image as an input. In this chapter, the task of multi-class tissue patch classification and segmentation is examined. Firstly, to gain a better understanding of how DCNNs classify histopathology images, their decision-making mechanisms are investigated with the visualisation of activation maps of different layers. Then, a multi-scale context-aware network is constructed to enhance the segmentation results for the ROI by incorporating contextual information to promote the discriminative power of the network. Following the processes considered above, the performances of different DCNNs with variant patch sizes and magnification levels are further compared. In the final section, the main efforts are contributed to achieve a pixel level segmentation result in a weakly supervised manner to further fine tune the classification outputs. In addition, to solve

the problem of under-training with insufficient annotation, a data synthesis/augmentation method is proposed to augment the training set for optimal model performance.

3.1 Overview and problem formulation

Histopathology images provide a valuable micro level inspection of a human or animal's tissue morphology which normally reflect the health state of the subjects. In practice, these images are captured through optical microscopes and digitalised by cameras attached to the scopes. With advancements in the last decade, scanners from different manufacturers with $40\times$ magnification capability are currently available in labs. One pixel in a $40\times$ image corresponds to around 0.25 mm^2 physical size in the slides. Hence, for a slide with a physical size of $5mm \times 10mm$, the captured image can easily reach 100,000px by 200,000px in terms of pixel size.

The extremely large resolution of a histopathology image makes it difficult to handle as a single unit with current computation capabilities. For most natural images, the problem can be addressed by introducing the resize operation, and then the semantic and abstract information of the objects can be maintained after the resizing operation. However, this does not work for histopathology images since the images are magnified on purpose to accommodate more details of the examined sample. As a consequence, the patch-based method is the most applicable and popular way to process WSIs as described in the literature. In fact, this is also the approach employed when scanning, storing, and visualising WSIs. The patch-based method divides a large image into small patches (tiles) for processing with a fixed size, and then merges the results from the small patches for a final output of the whole view. As illustrated in Fig. 3.1, with an original image with size $Hpx \times Wpx$, the actual processing units will be a much smaller region sampled from the original image with size $hpx \times wpx$ (where $H \gg h, W \gg w$) .

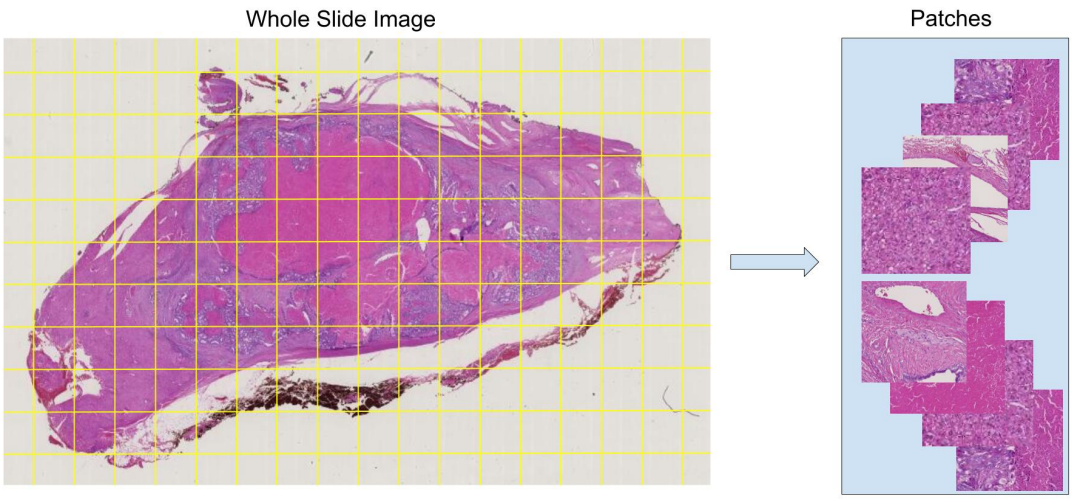


Fig. 3.1 From whole slide images to patches

The histopathology image can be processed with patches because they can be considered as a kind of orderless textural image with magnification levels greater than 10×. There are two categories of textures - repetitive and orderless. Fig. 3.2 shows the differences in textures between the repetitive texture and orderless texture. Based on the "Texton" theory [137], the textons in histology images correspond to the basic units and structures that exist in the tissue/cell, such as the nucleus and cell membrane. In contrast to artificial textures which often are repeated on a regular bias, the textons in histopathology images are randomly distributed and restricted to weak and latent rules.

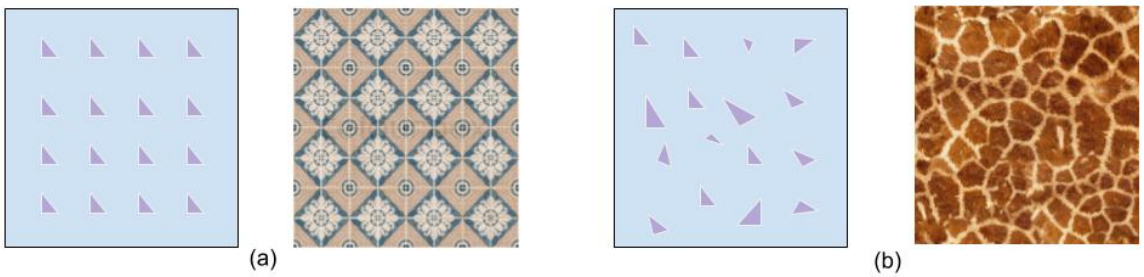


Fig. 3.2 Comparison between repetitive texture (a) and orderless texture (b)

Fig. 3.3 illustrates the components of histopathology images at a 40× magnification level from a "Texton" point of view. The elements including the nuclei, fibrosis, and blood cell

etc, in the third row, can be treated as the "Textons". A group of the same "Textons" will contribute the uniform texture features as shown in the second row. In the first row, different textures tangle together and form cases of mixed textures. Another important observation from Fig. 3.3 is the absence of any high-level abstract features in the images at the 40× magnification level. Only the "Textons" can be treated as objects with certain shapes and features. However, for the uniform textures or mixed textures, they are just a combination of the "Textons" with some statistical rules, and cannot be considered as objects in the analysis. Thus, the use of patches as the basic units for analysis is a practical solution, especially in the context of tissue classification and segmentation.

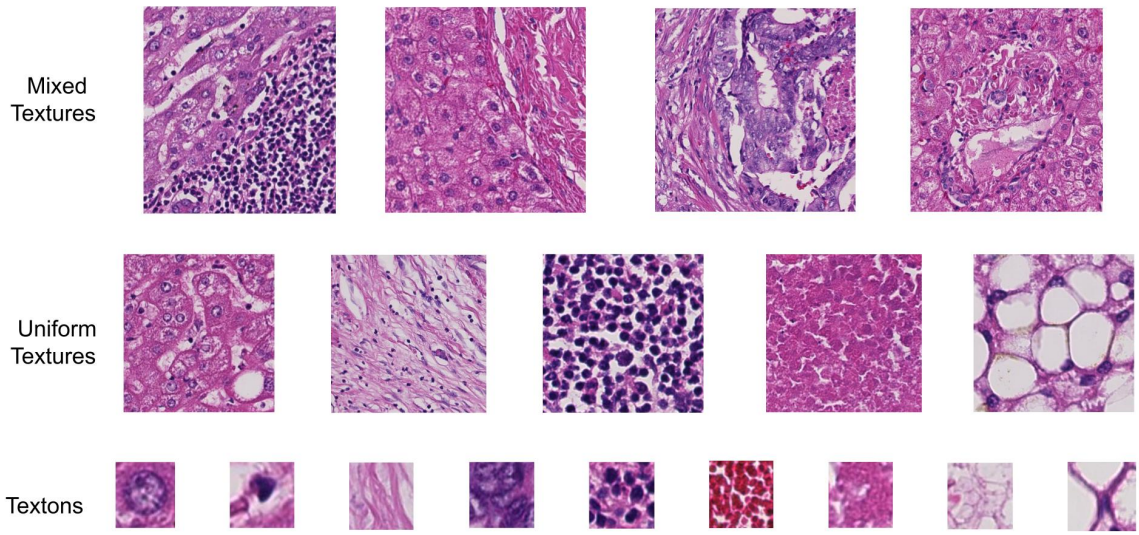


Fig. 3.3 Understand the WSI from the "Texton" point of view

3.2 DCNNs components and data preparation

Among different deep network structures, the DCNN is proved to be one of the most successful structure for image processing. This section starts by briefly introducing the convolutional neural network(CNN). Then the data collection and preparation process is described.

3.2.1 Components & structures of DCNNs

Convolutional layers A deep convolutional neural network consists a stack of different layers in which the output of the former layer is used as the input of the later one. For each input image I, the output of the nth convolutional layer y_n can be expressed as:

$$\begin{aligned} y_n &= C(I; w_1, w_2..., w_n) \\ &= C_n(C_{n-1}(...(C_2(C_1(I, w_1), w_2)...)w_{n-1}), w_n) \end{aligned} \tag{3.1}$$

where C_n represents the convolutional operation in the nth layer in the network, and w_n denotes the weights for that layer. The y_n are always referred to output of the nth layer. Normally, there are three kinds of operations inside one layer, convolution, pooling and non-linear activation. The convolutional layer is the core operation for a DCNN. The basic idea of a convolutional layer is to transform the image to a feature map with a series of filters (kernels) on a high dimension. With an image $I(x,y)$ and a filter $g(x,y)$ of the size of $w \times h$, the convolution operation is defined as:

$$(I * g)(x,y) = \sum_{v=y-h}^{y+h} \sum_{u=x-w}^{x+w} I(u,v)g(x-u, y-v) \tag{3.2}$$

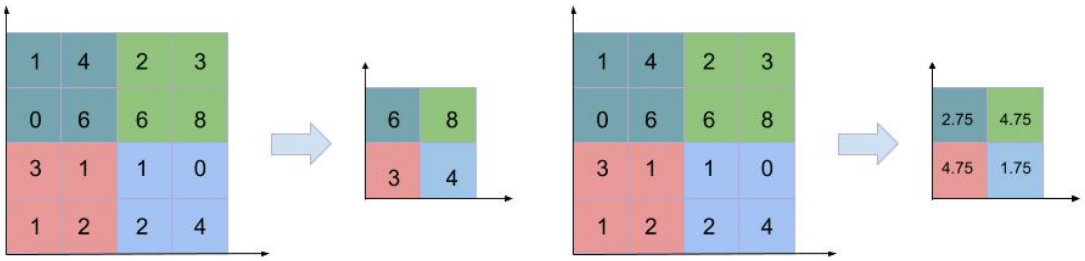
Where $*$ denotes the convolution operation. The outputs of the convolution will be linked to a local receptor field and then be mapped to a single value through an activation function. Non-linear activation aims to add non-linearity to the network. The rectified linear unit (ReLU) as defined in Eq. 3.3 has been proved to have better performance than the sigmoid function in DCNN:

$$\delta(x) = max(0,x) \tag{3.3}$$

One of the great advantages of CNN is that the number of parameters (weights) can reduce significantly via parameter sharing over images. After the training, the kernels can capture common spatial representations of the input images which can be clearly observed especially

from the first convolutional layer. In the first convolutional layer, the representations tend to extract the edge information which makes the kernels look like a Gabor filter or other first-order filters. For the kernels in the higher layer, the representations are not so obvious, since they contain embedded kernel information from lower layers.

Pooling can be treated as a kind of down-sampling operation. The idea is to divide the image into several pools (regions) and extract the maximum values or mean values to represent the region. Fig. 3.4 illustrate the pooling process with a 2×2 max-pooling and mean-pooling. Pooling operation is not mandatory for CNNs, but for the network with large



(a) Max Pooling (b) Mean Pooling

Fig. 3.4 Mean pooling and max pooling

input image size, it helps to reduce the spatial size of the feature maps as well as the number of the weights. Moreover, by summarising the information of the former layer, the pooling operation helps to transform the features to translation-invariant.

For the purpose of classification, the last layer of the network is often composed of a Softmax classifier layer. The output of the last convolutional layer or fully connected layer will be sent to the classifier. The Softmax function is the most widely adopted function for multi-label classification. The final label c is determined by the class j that has the maximum values $l(j)$ among the output vector.

$$l(j) = \frac{exp(z_j)}{\Sigma_{k=1}^{N} exp(z_k)}, for j = 1, ..., N \tag{3.4}$$

Dropout To prevent the over-fitting problem during training, drop-out is the essential trick to be added to the network. By adding the dropout parameters θ to a layer, the weights for the layer will be randomly set to zero with a probability of θ. Intuitively, this equals to adding noises to the images which will help to improve the model's generalisability.

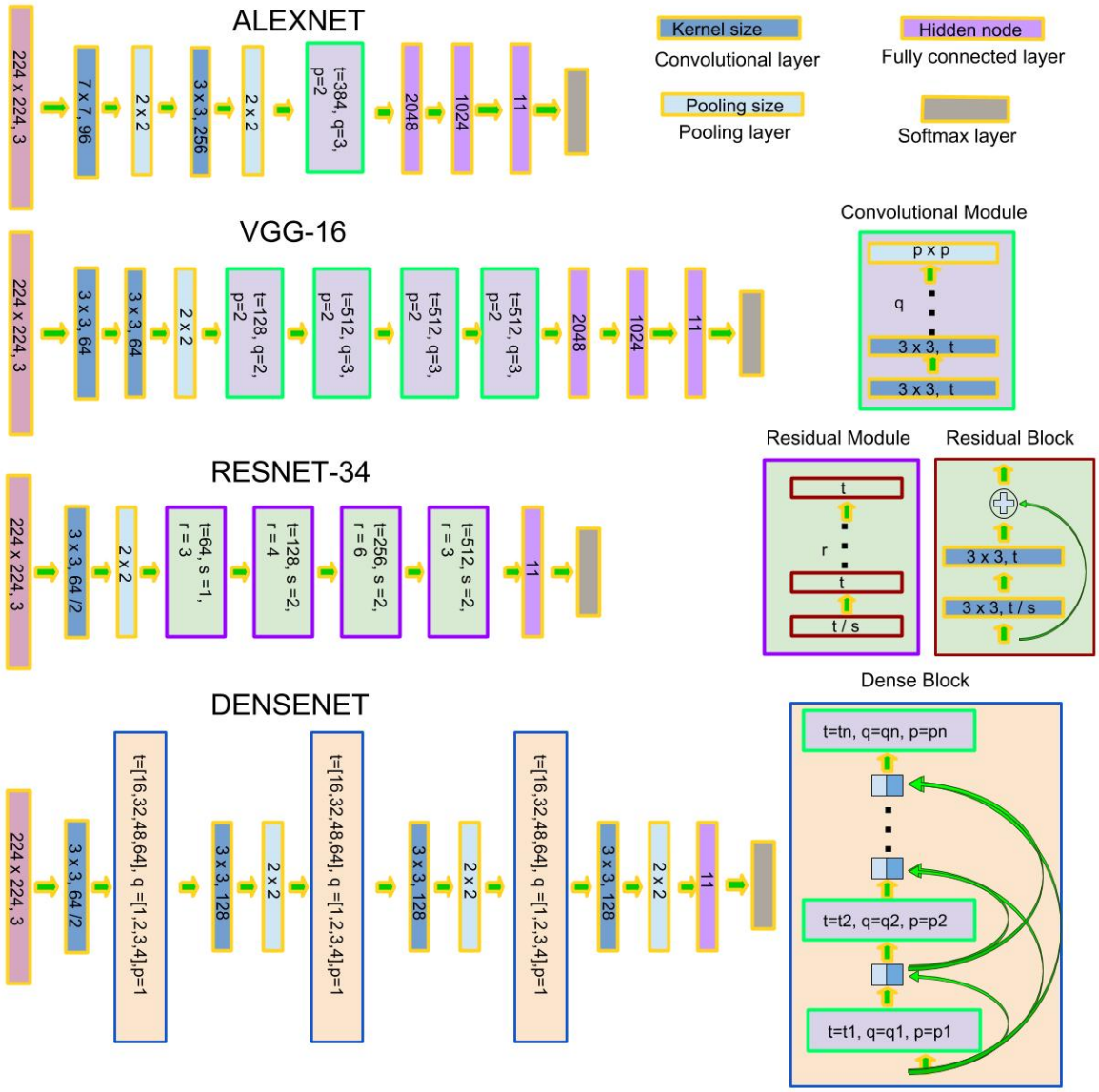


Fig. 3.5 Network structures of Alexnet, VGG-16 and Resnet-34

Batch Normalisation Batch normalisation is another import approach to enhance the performance of the network and accelerate the training process. It targets at normalising each of the

mini-batch during the training to overcome the gradient vanishing and exploding problem. During the training phase, the mean μ_B and the standard variance σ_B is calculated for each mini-batch. The output data y_B in the batch is normalised with two variables γ and β :

$$\hat{y}_B = \gamma \frac{y_B - \mu_B}{\sqrt{\sigma_B + \varepsilon}} + \beta \tag{3.5}$$

where ε is a small constant to prevent the division error, while, the γ and β need to be learned during the training. $\hat{y}_B$ is therefore the normalised data. In the test phase, especially when the size of mini-batch is 1, the μ_B and σ_B will be replaced by the mean and variance of the whole dataset.

Residual Block In some case when very deep networks are needed, the Resnet structure shown in Fig. 3.5, can be employed [138]. The shortcut in the residual block can be formulated as :

$$y = C_{1,2...n}(x, w_1, w_2 ... w_n) + x \tag{3.6}$$

where x, y is the input and output of the network, C is the convolutional layers. This kind of structure enables the network to pass the features from lower levels to a higher level to acquire more complicated features.

Dense Block Dense blocks are the main functioning modules for Densenet [139]. Instead of a skip layer element-wise addition as defined in Eq. 3.6, a dense block concatenates outputs from all the previous layers as its input. For the nth layer, the output can be defined as:

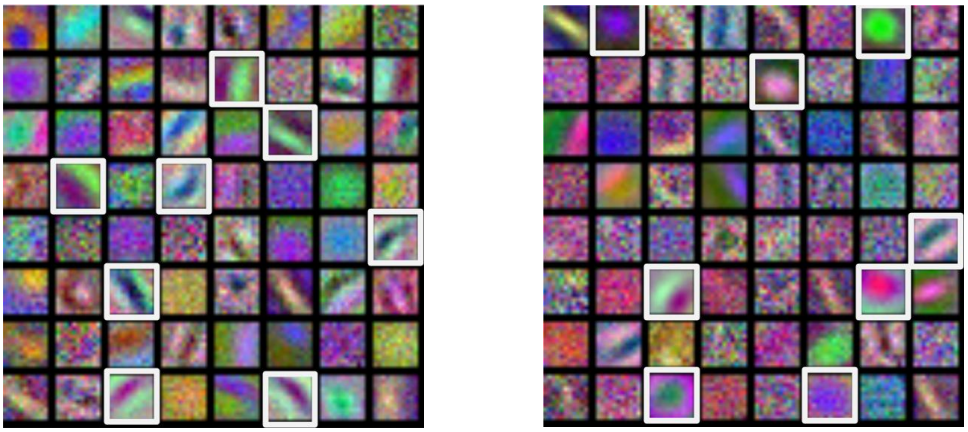
$$y_n = C_n((C_{n-1}(y_{n-2}, w_{n-1}), ..., C_2(y_1, w_2), C_1(x, w_1)), w_n) \tag{3.7}$$

It is reported that Densenet structure outperform the Resnet structure in multiple tasks.

Network Structures Four classic DCNN structures are discribed in the Fig. 3.5, Alexnet [140], VGG-16 [141], Resnet-34 and Densenet. Both of dropout and batch normalisation layers are facilitated for all the networks. The activation function employed here is ReLU.

3.2.2 Visualising and understanding the network

To understand how the CNN works for histopathology images, the features from different layers of the Alexnet are extracted and visualised through the deconvolution method proposed in [142].



(a) Features captured on 40× objective (b) Features captured on 20× objective

Fig. 3.6 Visualising first layer feature maps of models trained at different magnification levels. The highlighted boxes indicate in (a) correspond to the edge features of the nuclei and in (b) related to the shape features of the nuclei.

In Fig. 3.6, the features captured from the first convolutional layer of Alexnet illustrate that, compared to the network on 20× magnification level, the one on 40× magnification level lays more concentrate on the edges of the nuclei which means the network utilises more details from the nuclei or other components for the decision-making. While the features extracted from 20× magnification level (b) have more round blobs, which correspond to the outline of the nuclei. That means the morphologoies of the nuclei are processed as a whole. These observations coincide with the information that the input images can provide for human.

The feature maps from the 'conv5' layer are plotted in Fig. 3.7. To understand what kind of features matter to the network, the highest value in the map is deconvolved to locate and reconstruct the input. As shown in Fig. 3.7, for a patch comes from hepatocyte region, some of the feature maps try to capture the nuclei of the normal hepatocyte cells, which means

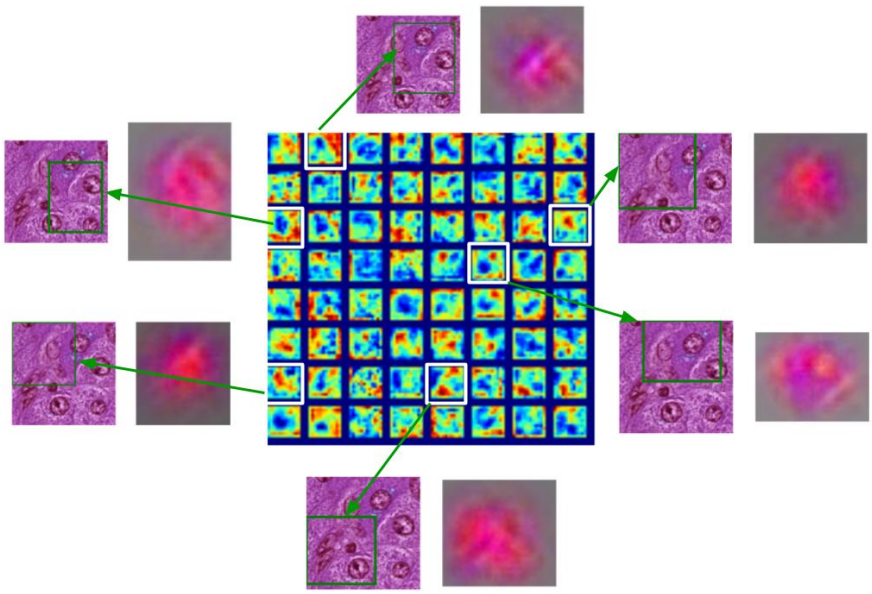
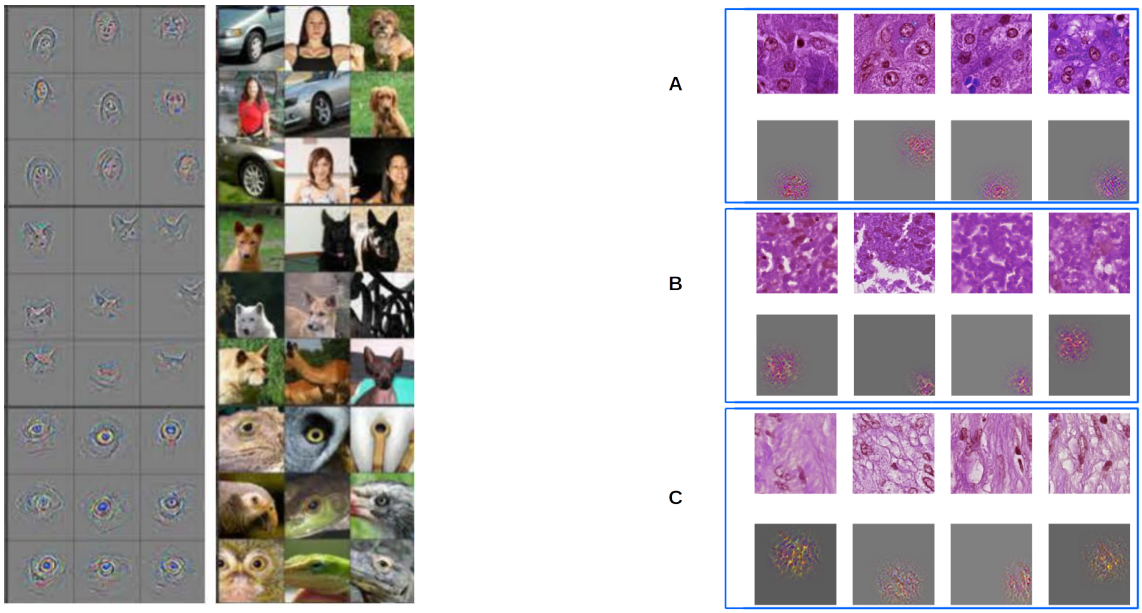


Fig. 3.7 Feature maps of the layer conv5 and the corresponding reconstruction of the top value in the feature map. The green bounding box defines the reception field of that top value.

the presence of these nuclei provides important evidence for final decision making. More examples are provided in Fig. 3.8. For the patches in the same group, the reactions are from the same feature map. From the group A in Fig. 3.8, it's easy to observe that the features captured for hepatocyte mainly are the nuclei's shape and edges, while for group B and C, the texture information are captured.

3.2.3 Dataset preparation and augmentation

Specimen preparation and digitalisatoin The WSIs are provided and annotated by the liver pathology experts in Karolinska University Hospital (KUH), Sweden. The dataset contains 104 WSIs, which are produced from resected specimens of different CLRM patients. After a series of processes like slicing, fixation and staining, the resected tissues are made into slides. Then the slides are digitalised with a Hamamatsu scanner. The source length of applied is 40$\times$ objective, and there are around 100,000 $\times$ 100,000 pixels in each WSI. The



(a) (b)

Fig. 3.8 Features of top reactions in the layer5 for different patches.
(a) Features reconstructed from layer5 for natural images from [142] (b) Features reconstructed from layer5 for CRLM histology images. Patches in A are from hepatocyte, B from necrosis, C from Fibrosis

corresponding physical size of each pixel is 226nm.

Dataset annotation process The annotation work is done manually by the experts in KUH with the software NDPviewer [1] which comes along with the scanner. Because of the large size and ambiguous details sometime, it is not possible to annotate the WSI slide-wise. Thus, the annotation work is done on the selected ROIs.

The annotation of this dataset was conducted in two phases. In the first phase, the annotations are mainly concentrated on the featured ROIs. Of all the 18 WSIs received, there are 30 ROIs selected and annotated in total. For each WSI, at least one region of interest is selected in each slide. While, the size of the ROI is not fixed, but they are sized at least 5000×5000 pixels in $40\times$ magnification objective. The overall workflow for the annotation in the first phase is described as follows:

[1] https://www.hamamatsu.com/eu/en/product/type/U12388-01/index.html

1. The experts go through the WSIs to choose the ROIs and box the regions.

2. In an ROI, each region that contains the same kind of tissue is annotated with an enclosed freehand boundary drawn by an expert. Then a tag will be used to denote the category of the region. Ten different kinds of tissues are annotated in the slides except for the background . They are *Hepatocyte*(H), *Tumour*(T), *Necrosis*(N), *Fibrosis*(F), *Inflammation*(I), *Blood*(B), *Foreign Blood Reaction*(FB), *Macrophage*(MF), *Mucin*(M) and *Bile Ducts* (BD).

3. In addition to the large ROIs, other regions which contain special types of tissues are annotated directly without the bounding box to keep the balance of the total samples.

4. Each annotation is extracted and categorised in terms of their annotation. Then the annotation will be reviewed by another expert to minimise the inter-observer subjectivity.

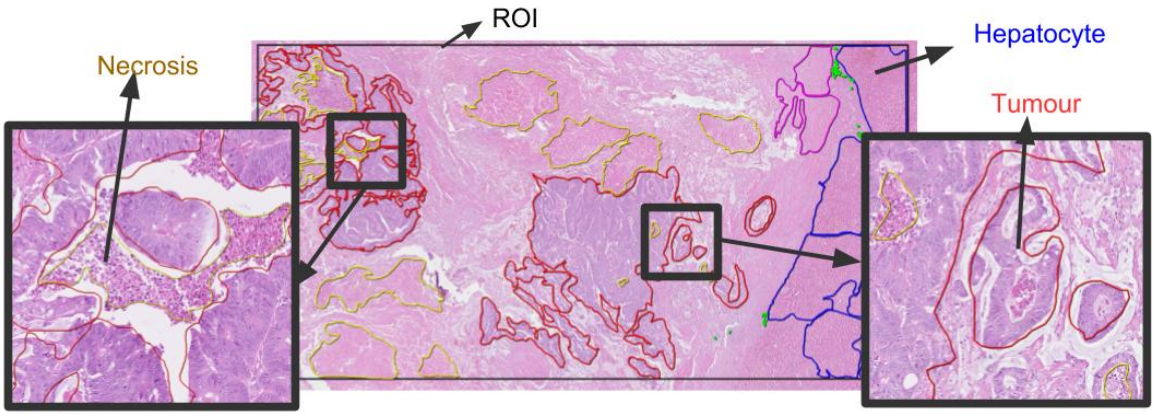


Fig. 3.9 Example of an annotated ROI from the CRLM dataset. The regions inside red contours are tumour, yellow contours are fibrosis, blue regions are hepatocyte.

An example of the annotation is demonstrated in Fig. 3.9. The ROI is marked with black rectangular bounding boxes. Within the ROI, each region of a uniform type of tissue is indicated by an enclosed irregular boundary drawn by free hand. Although the annotation boundaries are drawn with high quality and great efforts, inter- or intra-annotators variability

is still expected and observed largely due to the fact that the border between different tissues is often unclear. Hence, the principle employed during the annotation is to make sure the accuracy of the tissue inside the annotation area instead of the maximising the area or defining a precise border. For the test and evaluation purpose, the dataset are divided into two parts. 4 of the WSIs which only contains one ROI per slide are selected for evaluation. Out of the other 24 ROIs, 18 of them are used for training and 6 are used for evaluation.

In the second phase, another 31 slides are received. In these slides, the experts mainly annotate the areas with morphologies that do not appear or appears very few times in previous annotations. Of all the 49 annotated images, Table 3.1 provide the summary of the number of annotated tissue.

Table 3.1 Details of the annotated WSIs

	H	F	N	T	I	M	B	MF	FB	BD
No. Annontations	3148	2522	1112	6432	1253	428	296	443	176	1636
No. Patches	45102	12399	28092	22726	1827	10181	1030	270	2585	380

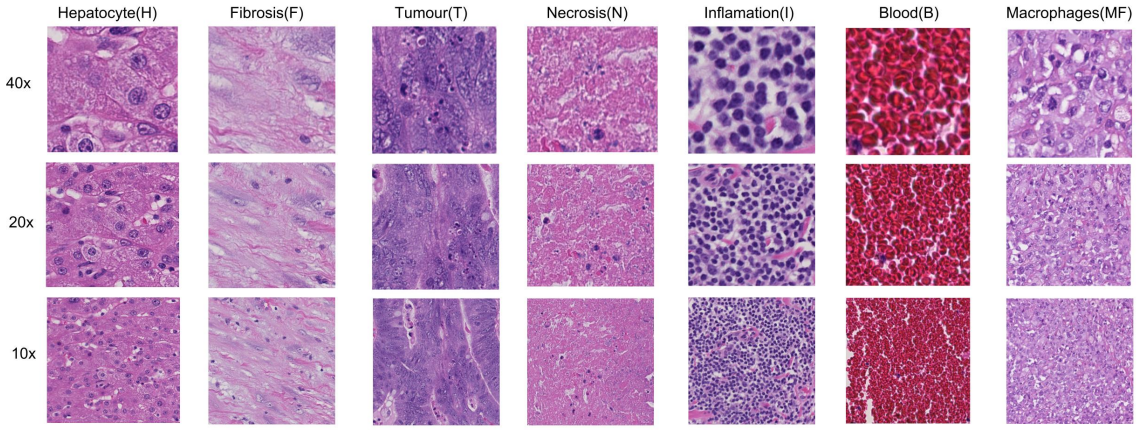


Fig. 3.10 Patches from different magnification levels and different classes

Patch cropping For each 128×128 patch within the annotation area, if more than 75% of the area is annotated as one tissue type, then this patch will be defined as a sample of

the corresponding category. At the same time, the surrounding pixels will be included in the sample as the context for classification. Thus, the size of a training sample can be very flexible. Whereas, in this study, to explore the effect of different objectives in this task, patches are prepared on $40\times$, $20\times$ and $10\times$ objectives with patch size $224 \times 224px$. Furthermore, patches with size of 448×448 are also prepared to analyse the importance of the contextual information. Due to the fact that the area of the annotation varies, an adaptive stride strategy is adopted to balance the number of different samples. The adaptive stride strategy can be formulated as Equation. 3.8:

$$stride = b + b \times min\{n \in Z | \log \frac{a}{p \times p \times 16} \leq n\} \tag{3.8}$$

Where a denotes the area of the annotated region, p is the patch size, b represents the basic stride, for $20\times, 40\times$ objective, it is 32, 64 accordingly. Fig. 3.10 illustrated the patches sampled from different objectives. On the $40\times$ objective, the patches contains details of the nuclei and the texture of the stroma and necrosis. Whereas, on the $20\times$ objective, the outline of the nuclei and the distribution of the cells can be clearly observed. On the $10\times$ objective, the outline of the nuclei is obscure and the detailed texture information of the stroma is lost as well.

Data augmentation Due to the limited number of annotated ROI, two augmentation approaches are adopted to enlarge the datasets and improve the robustness of the trained model.
1) Geographic transformation. For the WSI itself, there is no canonical orientation for the slide, but to the model, the orientation makes a difference. Hence, the original input images are flipped up to down and left to right to enlarge the database
2) Colour perturbation. The red channel and the blue channel of the input image are multiplied by a random value from 0.9 to 1.1 to change the colour. Since for HE-stained image, the concentrate of the dye or the thickness of the slide will lead to the differences in the red/blue ratio. Afterwards, an overall random factor in the range (0.8, 1.1) is applied to simulate the

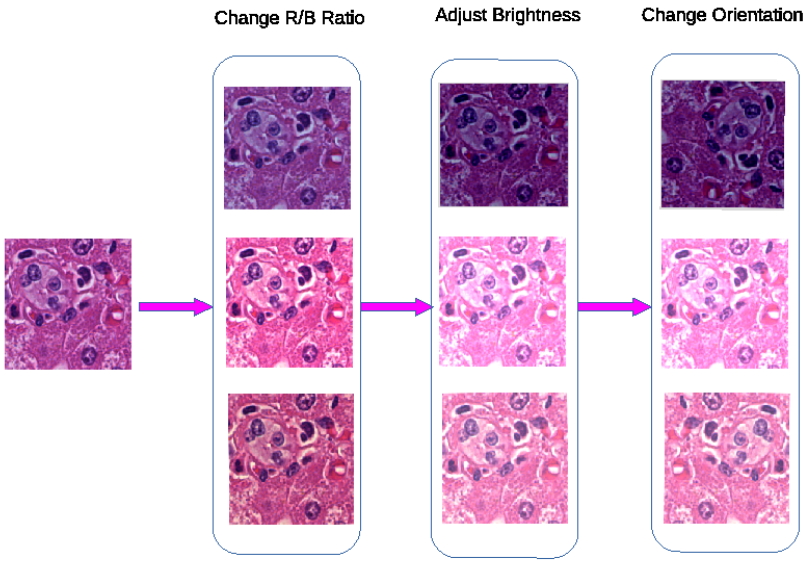


Fig. 3.11 Illustration of the data augmentation methods

change in brightness and contrast. At last, the pixel value of the image patches is normalised to the range (0, 1).

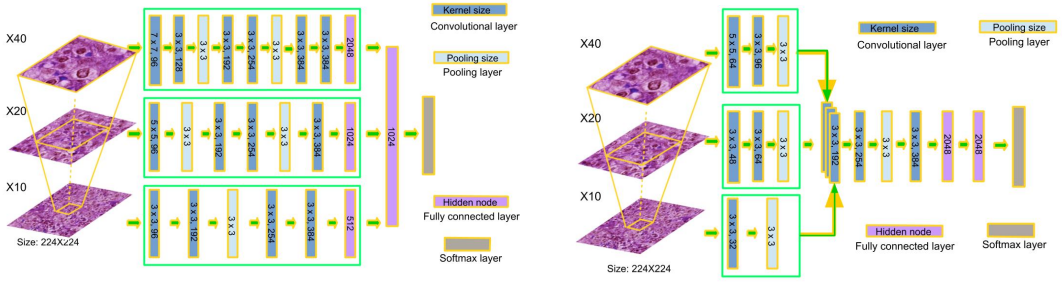
3.3 Multi-scale context-aware convolutional network (MCCNet)

For natural images, the top levels of the network tend to capture the abstract features of the input objects. Networks with more layers usually achieve a larger reception field which leads to better performance. However, histopathology images are very different from natural images. Firstly, there are no well-defined high-level abstract features that characterise the images. Thus, down-sampling does not help to differentiate the several tissue components. Even worse, it blurs the image and leads to loss of morphologically relevant low-level texture features. Secondly, because of the patch-based method, each patch is treated individually and loses contextual information. Therefore, in this section, two novel multi-scale context-

aware convolutional networks are proposed that are specifically designed for dealing with histopathology images, taking into account their unique visual characteristics. They are referred to as MCCNet1 and MCCNet2.

3.3.1 Network structure and implement details

The structures of the proposed network, MCCNet, are illustrated in Fig. 3.12. The input images to the networks are of the same size but from different magnification levels. The patches of higher level are extracted from the same region at different magnification levels sharing the same centre. Therefore the patches from a low magnification level provide contextual information of the region while the ones from high magnification level provide the details of the texture features. In MCCNet1, a late fusion strategy is adopted. The three sub-networks for different levels are trained separately before concatenating the feature vectors for prediction. For MCCNet2, an early fusion strategy takes the initiative. The feature maps are merged at an early stage and then trained together for the final output.



(a) MCCNet1 (b) MCCNet2

Fig. 3.12 Proposed network structure: MCCNet1 and MCCNet2

The network is implemented with the MXnet framework, while the training is carried on a dell desktop with the 2GB GPU (GTX-750) and 16GB memory. The dataset from the first annotation phase (18 WSI, 30 ROI) are used for training and testing. For the MCCNET1, the subnets are first trained separately. Then, with the trained weights, the network is fine-tuned with a relatively small learning rate (0.0001). For the MCCNET2, the network is trained

directly with the learning rate 0.01. The batch size hired is 32. For one training epoch, both networks take around 1.5 to 2 hours. For this training, all the nine categories annotated by the experts are involved. After 20 epochs of training, MCCNET1 converges with training accuracy 99.84% and validation accuracy 99.24%. For MCCNET2, it takes 30 epochs to converge, the training accuracy is 99.83% and validation accuracy 99.32%.

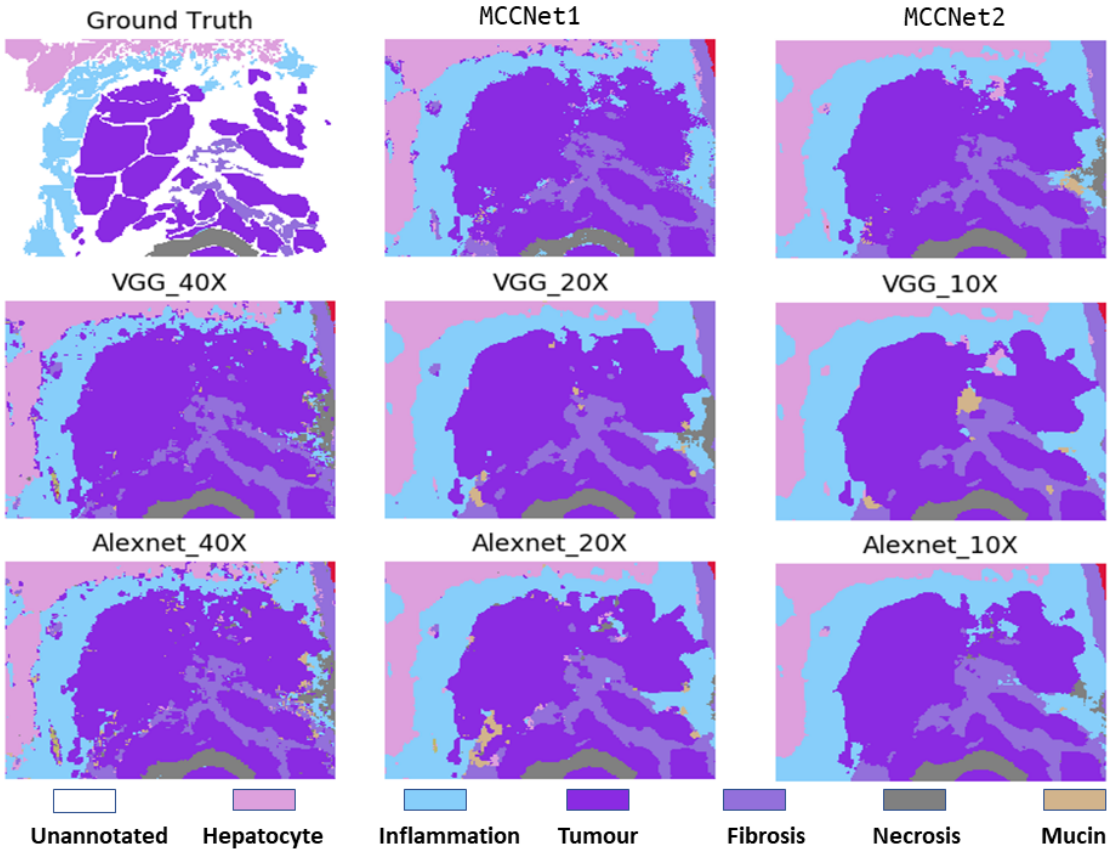


Fig. 3.13 Performances of proposed MCCNet on the ROIs

3.3.2 Results and discussion

The proposed multi-scale context-aware convolutional networks outperformed the classical methods Alexnet and VGG-16 with input from each single magnification level on the CRLM dataset both in terms of quantitative and qualitative results. In Table 3.2, the overall

classification results from the several networks and magnification levels are detailed. The results indicate that the patches from 20× magnification level outperform the other two levels and VGG-16 outperforms the Alexnet; MCCNet2 results in a slightly better score than the MCCNet1, but both outperform the start-of-the-art networks with single input magnification level. This indicates that the hierarchical input had a positive though the modest impact on the overall accuracy. However, the differences are more obviously appreciated in the qualitative results shown in Fig. 3.13. On 40× magnification level, the networks were very sensitive to variation due to high tissue details and lack of contextual information, and thus the results appear noisy. But on 10×, the patches contain more structural information of the tissue and fewer texture details, and hence it generates results closer to the semantic perception of the human eye.

Table 3.2 Overall quantitative classification accuracy of different models

Scale	Alexnet (%)	VGG-16 (%)	MCCNet1 (%)	MCCNet2 (%)
40×	94.39 ± 0.78	95.04 ± 0.59		
20×	95.97 ± 0.66	96.27 ± 0.39	96.43 ± 0.32	96.51 ± 0.35
10×	95.82 ± 0.52	96.20 ± 0.44		

Detailed examination of qualitative results in Fig. 3.13, it reveals that the main differences between network outputs lie within the unannotated areas, which are not included in the quantitative evaluation that is based on experts' annotations. That contributes to explain why the accuracies reported in Table 3.2 are very similar. To obtain an evaluation of the networks' performance in the unannotated regions, the experts are invited to visually assess and score the results. Again, both MCCNets lead to higher scores than the single input ones. Fig. 3.14 illustrates in detail the overall results of Table 3.2 by presenting the F1-score of the four main tissue classes of the dataset. The F1-scores of hepatocytes are much higher than for necrosis, fibrosis, and tumour cells. Meanwhile, the performances of different networks at several magnification levels are very close to each other. These results indicate that the hepatocytes in this dataset have the most distinguishable and stable features on all

magnification levels. However, for the necrosis and fibrosis classes, the proposed MCCNet methods achieve superior performance, which indicates that contextual information is more important for these classes.

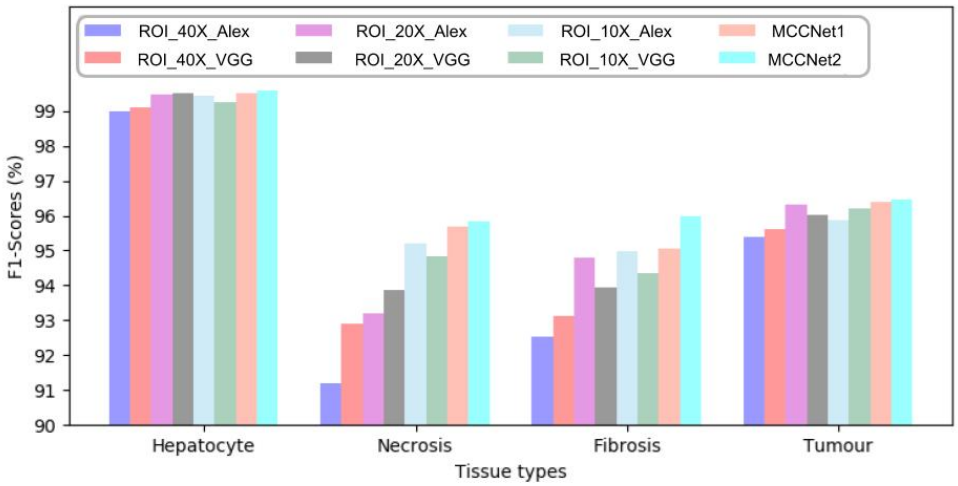


Fig. 3.14 Comparison of F1-scores on different classes

3.4 Whole slide image analysis using DCNNs

The objective of tissue patch classification to classify the patches by learning the underlying connections between the morphological features of patches from annotated regions and their corresponding annotation labels, and predict the labels for unannotated patches. As long as all the patches can be classified accurately, a rough segmentation can be acquired by stitching them back together. In most cases, the classification of a patch between cancerous or non-cancerous is the main concern, forming a binary classification problem. While in this study, more dedicated research simultaneously considering multiple classes is conducted. This is particularly useful in quantifying the proportion of different components of the tissue.

3.4.1 Methods and experiments

To explore the impact of networks with different depth to the classification performance and to find the most suitable network structure for the task, four different network structures, as illustrated in Fig. 3.5, with different depth, are employed in the task. Furthermore, to explore the importance of the contextual information or perception field, a dataset with different magnification levels ($40\times, 20\times, 10\times$) and different input sizes ($224 \times 224,\ 448 \times 448$) are investigated in the experiments.

The training of the deep models is performed based on the Pytorch framework and is running on TITANX-Pascal GPU. For all the networks, Stochastic Gradient Descent (SGD) is adopted for the back-propagation. The adaptive learning rate is used in all the networks with an initial learning rate 0.001, then decrease for every 1000 steps with the ratio 0.1. Dropout and batch normalisation are both adopted during the training.

The comparison of the performance in different models is conducted in two manners, quantitative evaluation and qualitative evaluation. Regarding the qualitative assessment, the main approach applied is to visualise the output vectors on a 2D space with t-SNE algorithm [143]. The t-SNE algorithm will map the high dimensional data to a 2D/3D space through the "Student's t-distribution" based on the distance of the two vectors. Furthermore, the whole slide segmentation results are calculated for both visual assessment and rough quantitative assessment. The quantitative evaluation is performed on two aspects, patch level evaluation and slide level evaluation. On the patch level , the accuracy on the testing dataset from CRLM dataset and an independent dataset from colorectal lymph node metastasis (CRCLNM) are calculated. For all the patches from CRLM dataset (as described in Table. 3.1) cropped out of the experts' annotations, 70% of them are used for training, 30% is used for testing. For the CRCLNM dataset, those classes that the same features as in the CRLM dataset, are used for further evaluation. On the slide level, the overall percentage of different components are calculated and compared with the ground truth information from IHC stained images.

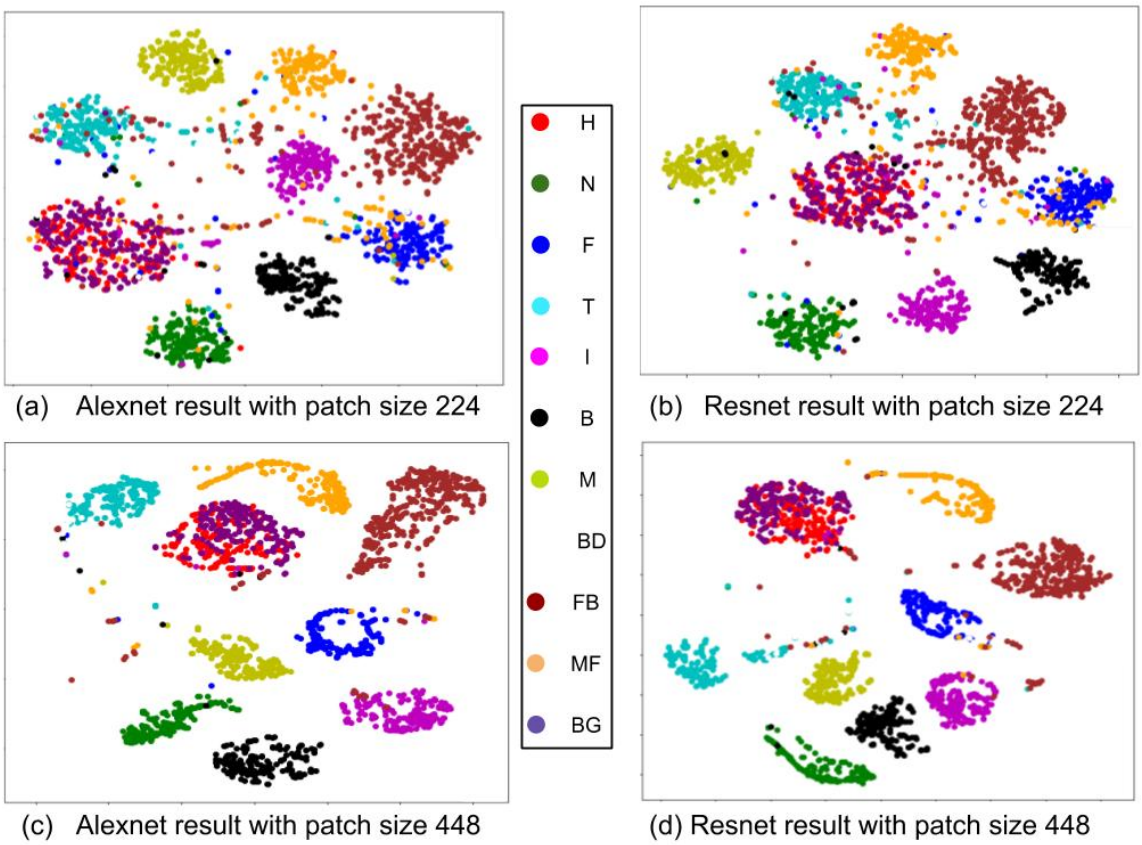


Fig. 3.15 Visualisation of feature vectors from the DCNN on 2d space with t-SNE

3.4.2 Results and discussion

Qualitative results In this part, two different approaches are presented to assess the classification performance visually. Firstly, the t-SNE algorithm is applied to visualise the 11 dimension vector output in a 2D space as shown in Fig. 3.15. The dots with different colours demonstrate patches from different classes. There are totally 2200 dots sampled from the testing dataset with 200 for each class. From the distribution, it is easy to find that all the classes can be well distinguished except for the classes, hepatocyte and background, tangled together. The reasons behind can be attributed to two reasons: one is because of the rough annotation of the hepatocyte classes, which include lots of background from porta area; the other one is due to the large white holes in the hepatocyte with fat that may cause confuses. However, this also provides us an index to evaluate the models' performance. In (c) and

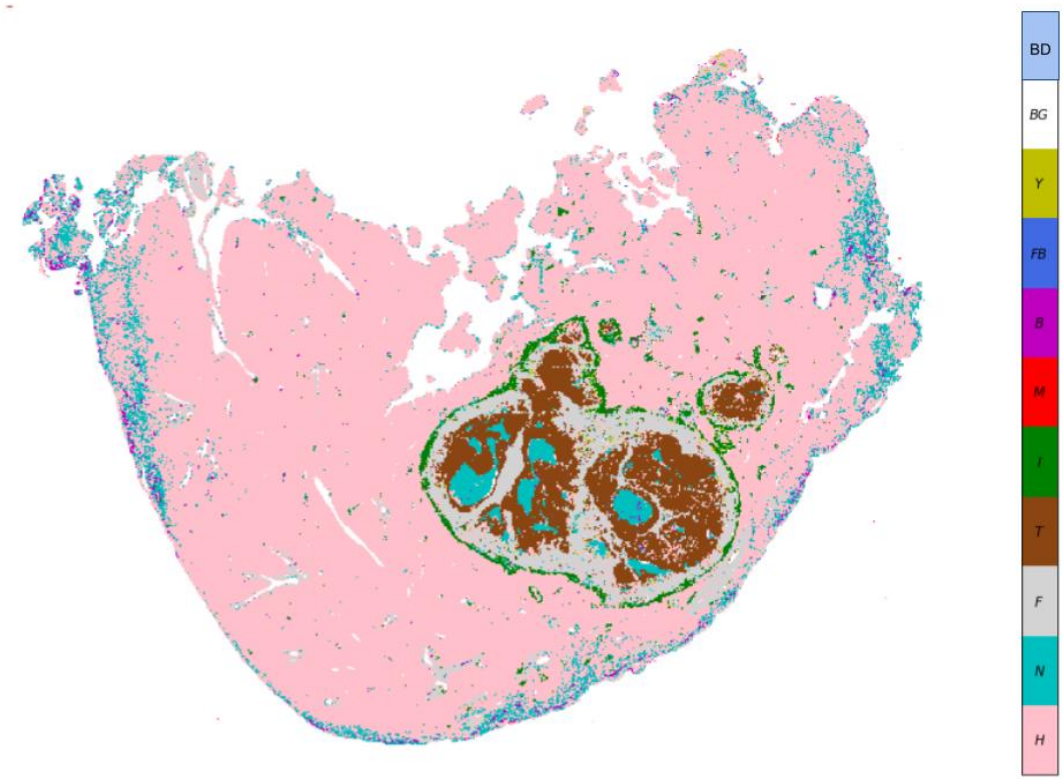
(d), it can be obviously observed that the two classes (hepatocyte and background) tend to separate from each other, while in (a) and (b), they are uniformly mixed. Observations from the distribution of the dots in Fig. 3.15 show that the models with patch size 448×448 have a very tight distribution, while the distributions for models with patch size 224×224 are relatively sparse. After comparing (a) / (b) and (a) / (c), it is natural to find that the patch size is a more important factor than the network structure.

Secondly, the trained model are applied on the WSIs in a sliding window manner. Thus, a segmentation result can be obtained as demonstrated in Fig. 3.16. With the first glance, the overall impression on the two results obtained with model Resnet-34 on two different patch size (224 vs. 448) is that there is more noise in the results by patch size 224. The reason for this phenomenon is obvious, the patch size 448 has a larger receptive field which can utilise contextual information for decision making to avoid misclassifying those small regions that have confusing appearances. However, the larger patch size will lead to a low resolution in the final result given the same input size. Especially on the borders where two kinds of tissue encountered. More results of WSIs obtained with patch size 448 are demonstrated in Fig. 3.17. It takes around 30 minutes to complete the prediction on the whole slide image with step size 112.

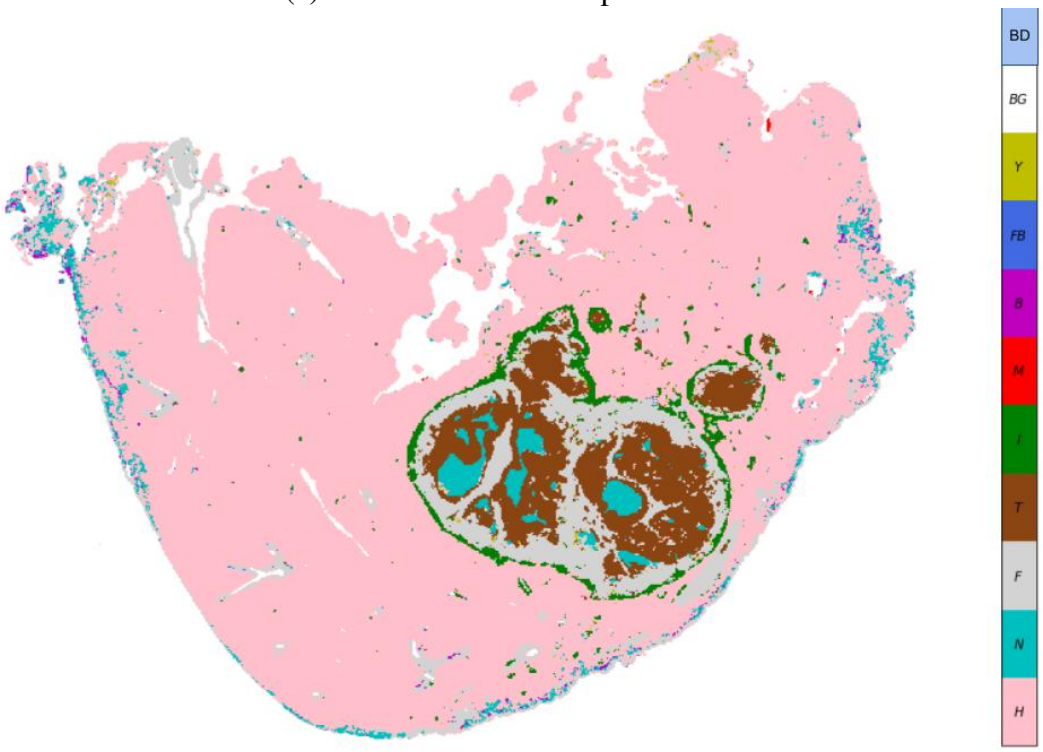
Quantitative results Table 3.3 listed the classification results on the testing dataset with different networks and inputs. Meanwhile, the independent test result with Alexnet, Vgg16 and Resnet-34 are illustrated in table 3.4. From the experiment results in Table 3.3, some basic conclusions can be drawn.

Table 3.3 Classification accuracy on the CRLM dataset

Objective	Patch size	Alexnet (%)	Vgg-16 (%)	Resnet-18 (%)	Resnet-34(%)	Resnet-50(%)	Densenet (%)
10×	224	92.02	92.53	92.24	93.73	94.13	93.93
20×	224	93.78	94.58	93.41	95.87	95.22	95.87
40×	224	93.55	93.27	94.36	96.01	95.19	94.85
40×	448	94.58	95.76	95.13	96.97	97.24	97.16



(a) Results on WSI with patch size 224



(b) Results on WSI with patch size 448

Fig. 3.16 Whole slide results with model Resnet-34

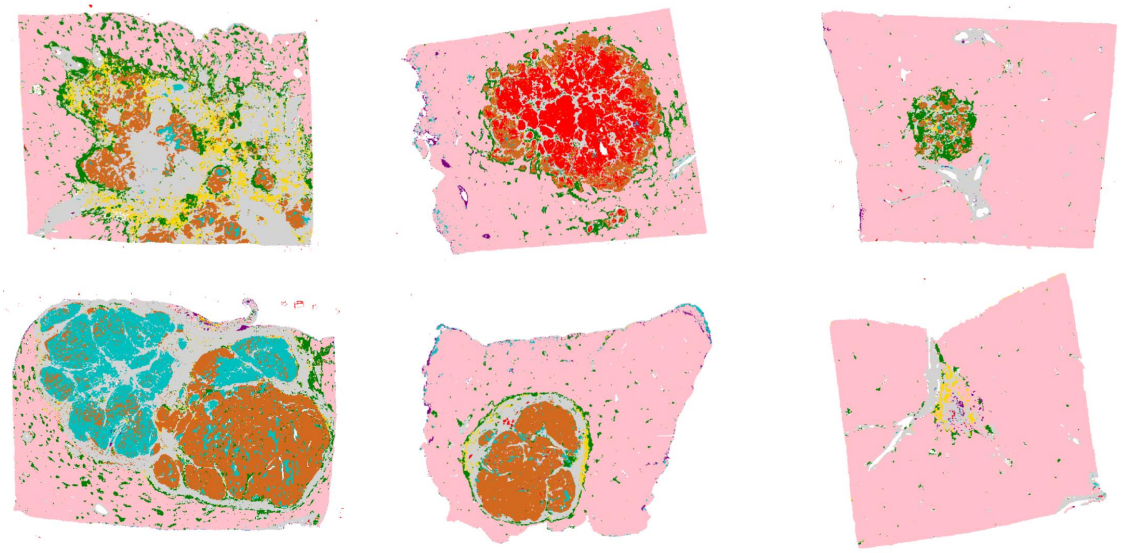


Fig. 3.17 Examples of results on WSIs with model Resnet-34 and patch size 448

(1) the magnification level or the patch size of the input plays a critical role in boosting the classification accuracy. As discussed in [58], it is difficult to choose the suitable patch dimension and magnification level. If the patch size is relatively small and the magnification objective is relatively high, the information will be limited for a tissue level classification which results in the decrease of the overall accuracy. On the contrary, the patches will contain too much information and lead to a bad performance in the whole slide image, because the larger patch size will not be able to describe the border of different tissue regions. With patch size 224×224 , the magnification level $20\times$ achieved a little better results than that on $40\times$, which indicate the reception filed(contextual information) is a very important factor for patches classification.

(2) The increase of the depth of networks will not necessarily help to improve the accuracy or generalisability. But with residual block, the increase of the network depth (Resnet-18, Resnet-34, Resnet-50) will indeed improve the performance to some extent. However, the deeper network also means a longer process time for the same amount of data, especially dealing with the WSIs.

(3) The features of different tissue types have different performance responses. For those tissue types with smaller scale "Textons" like fibrosis, necrosis and blood, the larger patch size will not help too much since the patches with smaller patch size already carry adequate information. Instead, in the process of classifying patches like hepatocyte and tumour, whose "Texton" is relatively large and sparsely distributed, the patch size will be more important.

Table 3.4 Evaluation accuracy on colorectal lymph node metastases

Model Name	Tumour (%)	Necrosis (%)	Fibrosis (%)	Mucin (%)	Blood (%)
Alex-224	78.48	81.74	90.13	86.76	**93.97**
Vgg16-224	87.22	69.86	90.13	90.21	93.21
Resnet-34-224	85.83	83.22	90.13	91.87	93.93
Alex-448	**92.08**	85.04	93.42	92.73	86.74
Vgg16-448	78.91	86.20	93.42	91.96	83.20
Resnet-34-448	90.69	**88.53**	**96.71**	**94.52**	88.64

A rough quantitative evaluation over the WSI is performed based on the IHC stained images which have high contrast over different types of tissue. The IHC staining employed for this task is CK18-CK19, which highlight the tumour & necrosis) as brown, hepatocyte as magenta and others as unstained. One example of the IHC/HE pair is demonstrated in Fig. 3.18. With simple processing, the IHC stained images (D) can be used as a mask as shown in (E). Since the WSIs are not exactly match, the evaluation can only be considered as a rough reflection of the accuracy of the segmentation. Hence, it is called as "pseudo ground truth", only the model Resnet-34 with patch size 448 is considered for the original input (A). The results of the model are illustrated in (B). The percentage of different kinds of tissues is demonstrated in (C) and (F) from the statistical value form (B) and (E).

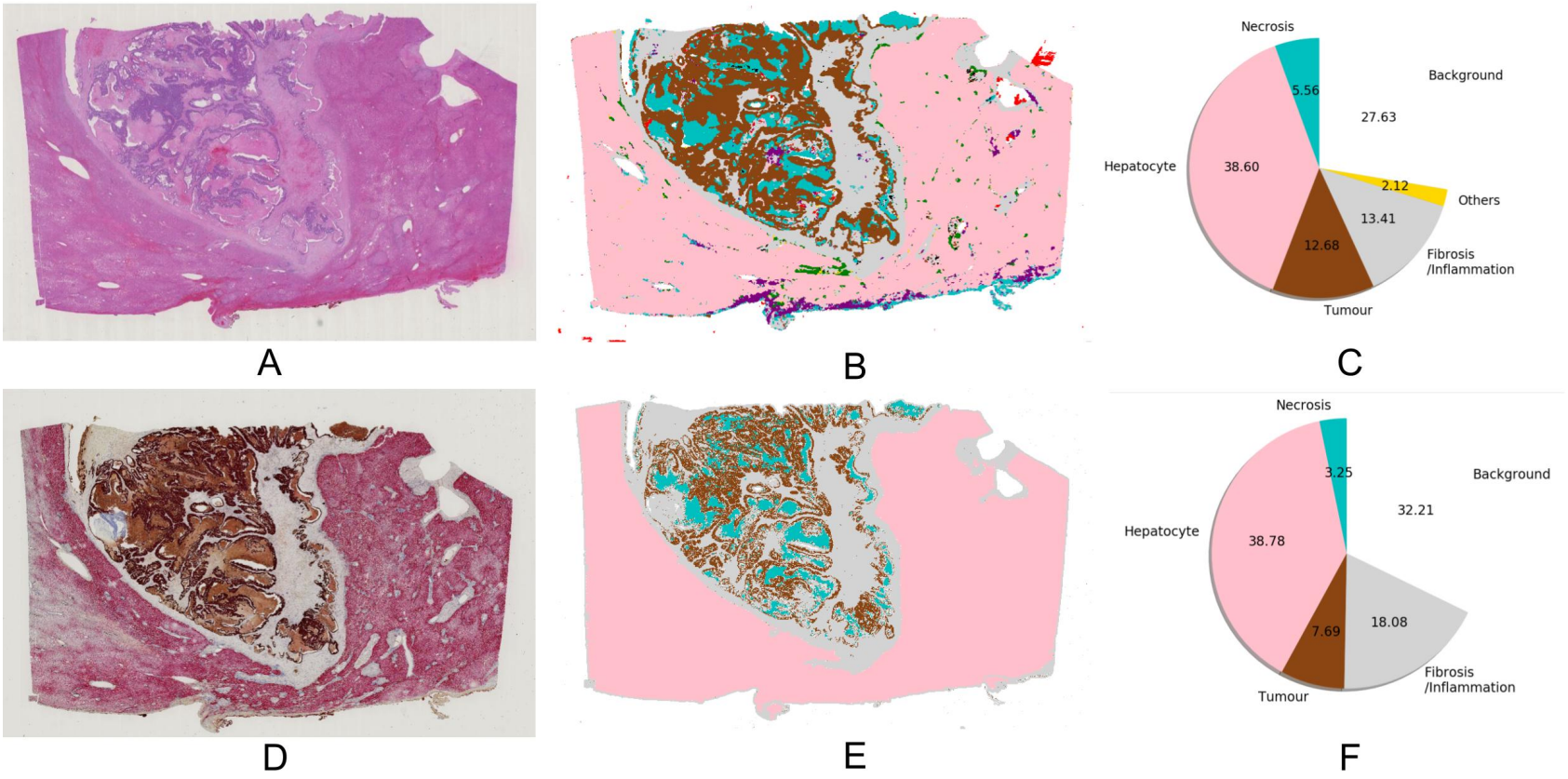


Fig. 3.18 “Pseudo ground truth” for whole slide segmentation evaluation. A: HE stained WSI; B: whole slide result from Resnet-34-448; C: quantitative analysis of the tissue components in A; D: IHC stained WSI; E: mask for the IHC stained WSI; F: quantitative analysis of the tissue components in C.

Through the observations, it is apparent the overall distribution of these two images are identical. Most of the differences lay on the percentage of tumour and necrosis which is because the model works on a higher level and will classify patches with tumour as tumour.

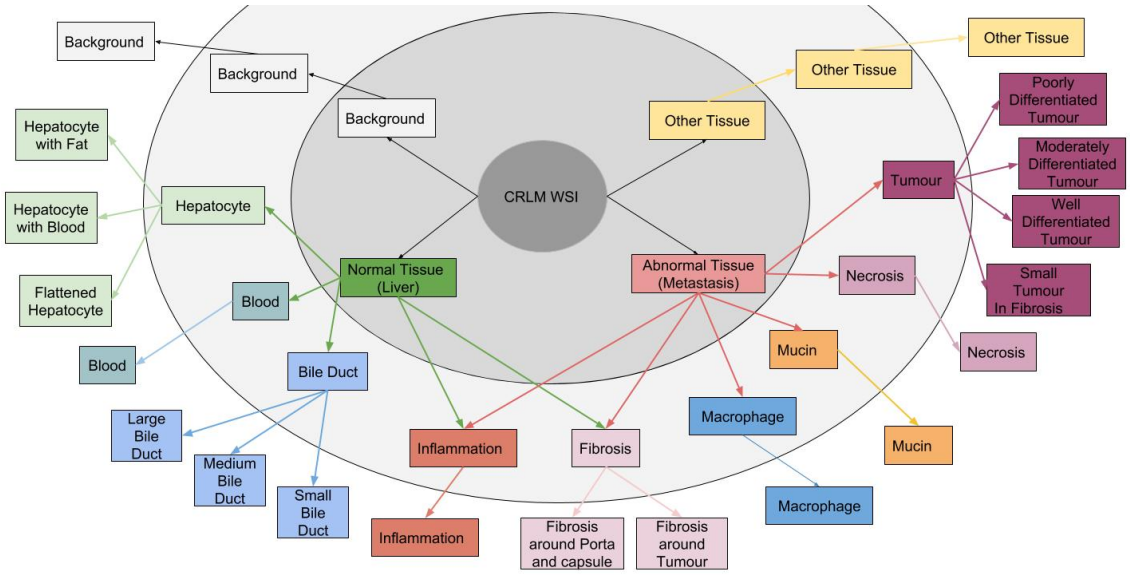


Fig. 3.19 Hierarchical labelling system for CRLM dataset

Further discussion By inspecting the distribution of the different categories on the t-SNE figure and the whole slide segmentation results. It can be concluded that a hierarchical labelling system (as illustrated in Fig. 3.19) with more tissue type is needed to take into consideration of more pathological definitions as well as the visual and morphological differences. From a higher abstract level or whole slide level, the tissue can be categorised as background, normal liver tissue, the abnormal tissue and other tissue. As the magnification level goes up, more details can be observed, while the classification criterion is also based on the component details. In this level, the tissue can be labelled with the nine classes mentioned above. However, this level of labelling can sometimes be too general. For example, fibrosis around tumour and fibrosis around capsule zone are totally different in term of their nature even though they may have very similar appearances. Vice versa, for the tumour class, tumour in different stages can look very different, but they are undoubtedly in the same class.

The outer layer of the diagram shows some more specific classes that can be considered in labelling and segmentation. According to the feedback from the pathology experts, except for the labelling system, the resolution and sensitivity of the segmentation output need to be further enhanced as well. Therefore, in the following section and chapters, variant methods are proposed to enhance the segmentation results based on the current annotations.

3.5 From tissue level to pixel level

In clinical related research, one of the prior requirements is to ensure the accuracy and confidence of the measured data. Even though the classification-based method described in last section can achieve decent results, it is still not accurate enough for a quantitative assessment. The ideal result of the tissue components analysis on WSIs is to produce pixel level segmentation on the input images. However, to achieve that goal, the main challenge to deal with is the insufficient and inadequate annotation data. Therefore, in the section, a weakly supervised method that integrate segmentation-based method with the fully convolutional neural network is proposed to further enhance the segmentation results in a hierarchical way.

3.5.1 Framework and methodology

Increasing the resolution of the output result is critical to achieve a more accurate overall result on the WSIs. However, there is no need to apply a pixel model on all the patches in the whole slide. Therefore, a two-step (border localisation and border enhancement) framework is proposed to enhance the resolution and accuracy of the WSI result for the purpose of reducing the computational cost over the WSI and preventing the over-segment happening.

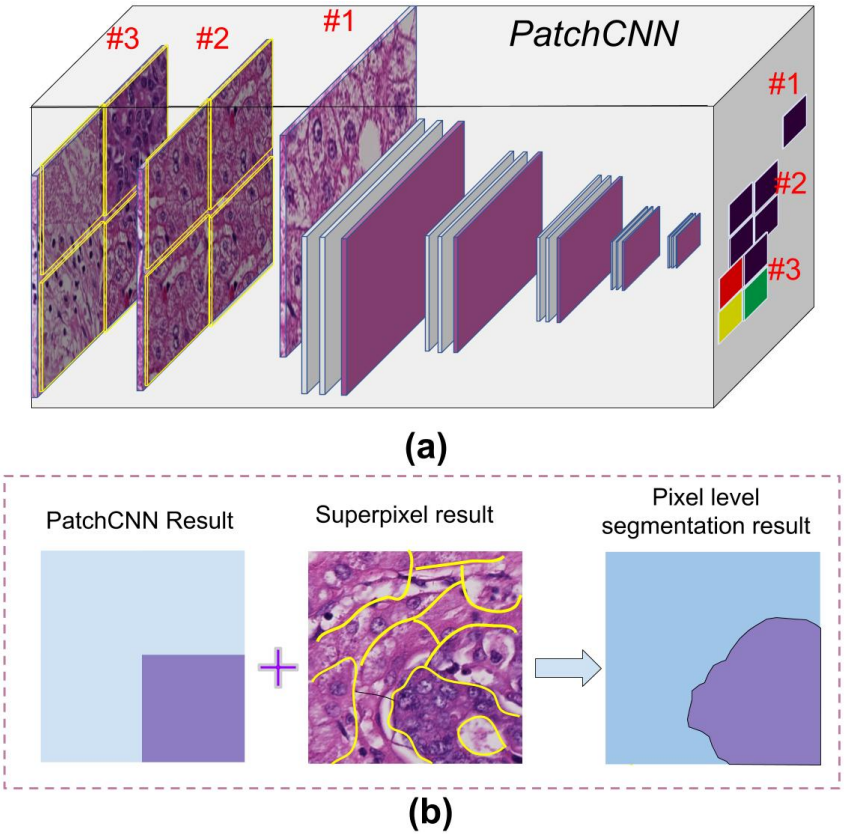


Fig. 3.20 Proposed framework for segmentation-based segmentation with border localisation (a) and border enhancement (b)

Border localisation : weakly-supervised PatchCNN framework

Data imbalance is one of the most challenging problems for machine learning. When it comes to segmentation problem, it is especially true since most of the patches are localised in homogeneous regions that do not need further process, comparing to those that fall in the tissue border region. Therefore, the first step of the proposed framework is to locate the heterogeneous region by determining whether the patches contain homogeneous regions or not. To this end, PacthCNN is proposed as illustrated in Fig. 3.20(a), which is actually composed of a fully convolutional neural network. Instead of generating one label, PatchCNN can output a $n \times n$ matrix to describe the content of patches while increasing the resolution of the results. The output will grow accordingly when the size increases. In this case, the

computation cost will be greatly reduced compared to using a sliding window method for getting output with the same size.

Even though the output resolution increases, the annotation did not. If the ground truth label of patches is the same for all the components in the matrix, there is no way that the model can model the inner spatial relationship of different types of tissues inside. For this reason, a weakly supervised method is introduced. The input images are divided into $n \times n$ regions (for example 2×2 in Fig. 3.20 (a)), each region will be filled with the cropped centre of a random patch in the database. Meanwhile, the ground truth labels for the output of the network will be changed. Thus, the network will be enhanced and the spatial resolution will be improved without sacrificing contextual information. With the trained model, patches with homogeneous tissue type and with mixed tissue type can be classified easily without annotating extra dataset. For instance, with n equal to 4, if all these 4×4 outputs have the same label and very high confidence (>0.9), then the patches can be treated as a homogeneous patch. Vice versa, those have different labels and low confidence of the labels are mixed regions.

Border enhancement: towards a pixel level results

As stated previously, unlike the natural image processing, the annotation of histopathology images requires the professional knowledge on the pathology. Hence the lack of pixel level annotation is one of the biggest challenges. In this circumstance, another weakly supervised method is proposed to enhance the final segmentation results. To achieve a pixel level result based on the PatchCNN result without ground truth on the pixel level , unsupervised superpixel algorithm [144] is employed to further enhance the results based on the intensity features of the input image. The workflow involving superpixel is demonstrated in Fig. 3.20 (b). The superpixel algorithm actually takes advantages of the local properties of the pixels and help to find the natural borders in terms of the derivatives of neighbouring pixels. The

initial points will be defined by the $2*n*2*n$, where n is the number of divisions of the patches. Afterwards, the pixels of the input patch will cluster into several blocks according to their intensity features. With the outputs of the superpixel and PatchCNN, it is natural to assign the regions with the classification labels with simple rules: the region will be labelled by the class that has the largest amount in terms of quantity of pixels in the region. After the fusion of these two results, the final segmentation output will become smooth and get rid of the rigid boundary.

3.5.2 Experiment details

Training and testing dataset Except for the original data, two kinds of synthesised data are employed to boost the model's performance. As shown in Fig. 3.21, the uniform synthesis (a) will make sure all the contents in the slides are consistent, while the random synthesis (b) attempt to make the different parts distinctive.

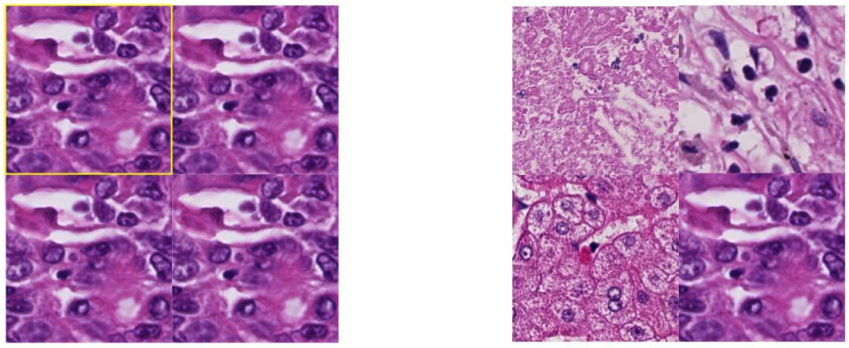


Fig. 3.21 Simple data synthesis strategy

Evaluation dataset The dataset for evaluating the performance of proposed methods is composed of 50 ROIs with pixel level annotation. The ROIs are cropped mostly from the tumour border area or metastasis area which has multiple tissue types inside with size 4096×4096. Afterwards, the ROIs are annotated by several experts. The final annotation is actually

an average over multiple experts. Few examples of these ROIs and their corresponding masks are demonstrated in Fig. 3.22 A and B.

Implementation details The PatchCNN model is constructed in a fully convolutional manner, which means there is no fully connected layers used in the network. The backbone network for PatchCNN employed is Resnet-34. Instead of attaching one fully connected layer to the end, two convolutional layers are added to extract the features for classification. One important trick for the PatchCNN is to replace the zero padding with the reflective padding to ensure a continuous feature and identical performance with larger input. During the training, for an input image with size 448×448, the output is 4×4, which lead to a distinguishable size 112×112. However, the corresponding reception field defined in the model is 128×128.

Evaluation metrics The evaluation of the performance of different networks is performed in two aspects. In terms of numerical and quantitative evaluation, the MIoU and PA value are hired for the comparison among different methods. While visual performance can be assessed through the results demonstrated in Fig. 3.22.

3.5.3 Results and discussion

To evaluate the performance of the proposed framework, two baseline methods are employed. Both of them have the same backbone network, Resnet-34, and obtain the segmentation results using the sliding window method. The difference lies if the post-processing method is applied. The Conditional Random Filed (CRF) introduced in [47] for natural scene semantic segmentation are employed to smooth the segmentation results. The MIoU values of different models are demonstrated in Table 3.5. For the purposed of qualitative assessment, output from different models is detailed in Fig. 3.22. As can be observed from the Table 3.5 that the proposed framework achieve the best result, even though the improvement is limited in terms of numerical performance. However, the final improvement comes from two parts, the improvement of the original network output (PatchCNN v.s Resnet-34) and the improvement

Table 3.5 Segmentation results (PA & MIoU) on the evaluation dataset with different models

	Resnet-34	Resnet-34 + CRF	PatchCNN	PatchCNN + CRF	PatchCNN + Superpixel
PA	77.31%	80.47%	78.12%	80.27%	80.87%
MIoU	57.92 %	59.33%	58.16%	58.78%	59.74%

of the post-processing (CRF v.s superpixel). From the visual assessment in Fig. 3.22, it can be concluded that both of the post-processing method, CRF and the superpixel, will help to enhance the semantic segmentation performance by removing the artefact and locating the real border. However, the difference from Resnet-34 and PatchCNN are very limited, which can not be observed visually, for example, the testing images on the first & second row in Fig. 3.22, different approaches have very similar outputs. When it comes to those heterogenous regions, for example, the testing images on the third & fourth row in Fig.3.22 with many small tumour areas, the performances become different. In summary, comparing the current results with the ground truth, there is a large space for the models to improve the segmentation results.

3.6 Summary

In this chapter, the patch-based method for histopathology image analysis is investigated. Attempts have been made step by step to achieve better segmentation results based on the patch-level annotations. Firstly, based on the understanding of both the challenges and how the networks make decisions from the experiment results, a multi-scale context-aware network structure is proposed to improve the segmentation performance through the combined information from different magnification level and enhanced the contextual information. Furthermore, different network structures are compared on the classification task to investigate the influence of different magnification levels, input sizes and network depth. In the end, efforts are made to improve the segmentation result in a pixel level with a segmentation-based

method in a weakly supervised manner. Even though there are improvements shown on both quantitative result and qualitative result, it is far from satisfactory. Hence, in the next chapter, cell level information will be employed for segmentation result enhancement.

Fig. 3.22 Examples of the segmentation results. A : original input image; B: ground truth mask; C: Resnet-34-448 ; D: PatchCNN ; E: Resnet-34 + CRF; F: PatchCNN + Superpixel

Chapter 4

Nuclei Segmentation and Classification for Cell Level Histopathology Image Analysis

Nuclei (cells) detection and segmentation are fundamental tasks for many biological or clinical related applications. The shapes, quantity and the distribution of the nuclei (cells) are distinctive features for experimental evaluation or decision-making. In many biology studies, counting the nuclei (cell) is an indispensable daily routine. Whereas, in clinical practice, the successful detection and classification of different nuclei (cells) can be considered as valuable evidence in assessing the progression of certain diseases.

In this chapter, the cell level analysis of this research is presented, addressing the CRLM task from three different aspects. Firstly, in section 4.2, with the help of image processing methods, nuclei are extracted from liver tissue and then modelled with a variational auto-encoder (VAE). Then the classification of different nuclei types is explored in a weakly supervised manner to overcome the issue of insufficient annotations. Secondly, a novel and efficient network, referred to as US-net, is proposed to deal with the nuclei detection and segmentation problem in tissue with dense or overlapped nuclei. In section 4.4, based on the

cell level results in this chapter, a tissue region growing algorithm is proposed to enhance the tissue level segmentation results in Chapter 3, which is the final objective of the cell level research in this chapter.

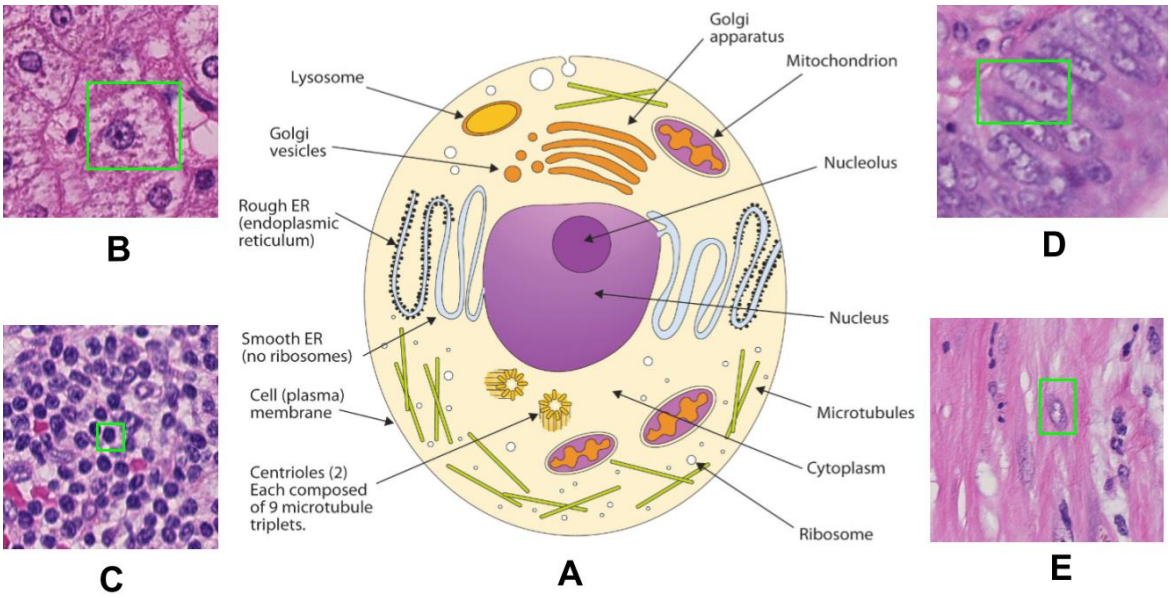


Fig. 4.1 A standard cell structure and examples of nuclei in the CRLM dataset (a) the sketch of a conventional cell structure, (b) nucleus of hepatocyte , (c) nuclei of inflammation cells, (d) nuclei of tumour cells, (e) nuclei of fibrosis cells

4.1 Overview

Cells are the elementary units of many living creatures, but are not solid objects. In fact, they are made of several components as illustrated in Fig. 4.1(A). The components for different cells can vary largely in terms of their functionalities in tissue. However, no matter how a cell looks, the nucleus, membrane, and cytoplasm are its indispensable components.

For histopathology image analysis in clinical practice, HE staining is the most common and economical way to highlight the basic tissue structure for better visualisation and examination. However, in HE stained histopathology images, only the nuclei (sometimes the nucleolus) can be clearly observed as meaningful individual objects from other components

in the background. Hence, nuclei related analysis such as detection, classification and segmentation are the main tasks at the cell level. One of the main challenges in manual traditional pathology evaluation based on HE stained slides, is the significant time, efforts and skills required for the visual assessment of each case. A massive number of samples are produced on a daily basis that need to be examined. Moreover, the increasing shortage of sub-specialised pathologists is also being reported [145]. Fortunately, with the recent advances in digitisation techniques for scanning digital whole slide images, a good foundation has been laid for developing intelligent computer-aided histopathology assessment systems. Such systems are expected to augment pathologists' ability by automating some of the fundamental, labour-intensive and relatively easy tasks, thereby allowing experts to focus on the most challenging parts of the assessment. The analysis of the cell shape, size, distribution and other features is an essential task for both biologists and histopathologists in their visual analysis of histopathology data. The automation of this task plays a critical role in the subsequent analysis in computer-aided histopathology image assessment.

Fig. 4.1(B-E) demonstrate a few nuclei examples extracted from the HE stained histopathology images of CRLM. It can be observed that nuclei from different types of tissues have significantly different morphologies. However, not all tissue contains cell structures, for instance, necrosis and mucin. While others have several different subtypes inside, e.g., for normal liver tissue (marked as hepatocyte) have hepatocytes, kupffer cells, *et. al.*.

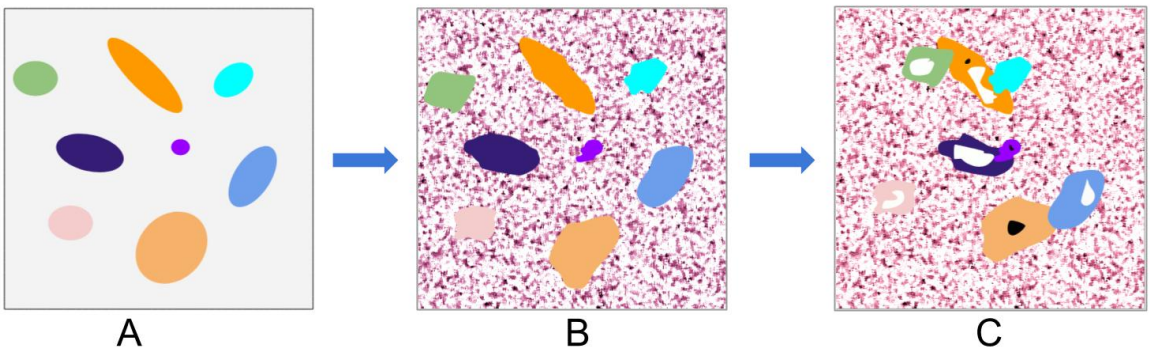


Fig. 4.2 Cell level problem statement and task overview

The nuclei related tasks in histopathology image analysis can be understood in a rather abstract manner. Fig. 4.2 illustrates the complicated nature of nuclei morphology. Fig. 4.2 (A) shows the ideal scenario where the cells/nuclei objects can be represented with regular shapes which indicate the different shapes and scales of these objects. The multi-scale problem is one of the many challenges that needs to overcome in this task. Whereas, in real histopathology images, there will be a lot of noise in both the background and foreground as shown in Fig. 4.2 (B). The outlines of the objects are mostly coarse instead of being smooth. A more realistic and challenging situation is described in Fig. 4.2 (C). For example, there are more textures to be considered in both the nuclei/cells and the background. Moreover, situations like overlay and crowding impose further challenges.

To analyse HE stained images, accurate nuclei detection is the basis for many tasks. The idea of nuclei detection is to locate the approximate position and size of the nuclei, *e.g.*, with bounding boxes. Sometimes, detection can be achieved with the semantic segmentation permitted under the condition that the nuclei are distributed sparsely. However, when it comes to a situation when the nuclei significantly overlap, nuclei detection is indispensable. Nuclei classification classifies the nuclei regarding their morphology and surroundings after the nuclei have been detected, especially in cases where different types of cells mix together. Whereas, the nuclei semantic segmentation task segments the nuclei from other components. This is not as relevant here due to the limited application scenario, but can help to improve the performance of the detection task. Nuclei instance segmentation which aims to separate the overlapped objects, is an approach that jointly addresses the nuclei detection, classification and segmentation tasks. In the output of the instance segmentation, almost all visual related features can be extracted and calculated, for example, the perimeter and the size of the nuclei. Nuclei instance segmentation is a promising solution to many nuclei/cell related analysis problems if it is appropriately designed. Thus, one of the main goals of this chapter to develop an effective nuclei instance segmentation approach.

4.2 Nuclei detection and classification in liver tissue - a weakly supervised method

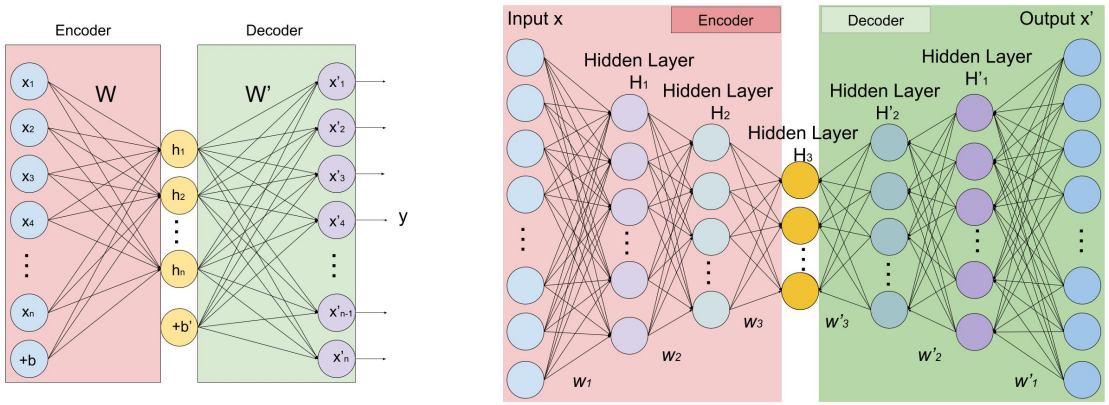
In supervised machine learning, adequate labelled training datasets are normally associated with the excellent performance of a trained model. However, when the available training dataset is inadequate, the performance of the model will drop significantly. To deal with the numerous unannotated data generated every day from clinical practice or biomedical experiments, unsupervised or weakly supervised methods provide practical alternatives with great potential. In this section, the traditional image processing techniques and unsupervised deep learning methods are integrated to investigate the problem of nuclei detection and classification with weak supervision.

4.2.1 Preliminaries

Auto-encoder is one of the most widely used unsupervised deep learning algorithms to model the representation of the data. According to the implementation methods and different constraints set to the network, several variations are proposed based on the basic auto-encoder, like sparse auto-encoder, denoise auto-encoder and variational auto-encoder. In this section, a few auto-encoders employed in this research are briefly introduced.

Basic Auto-encoder Auto-encoders usually includes two main parts in one network, the encoder and the decoder (pink zone and green zone in Fig. 4.3). The target of an auto-encoder is to compress the input into a feature vector with encoder and reconstruct the input with decoder with a minimum reconstruction error. For an input x, the feature vectors or, in other words, the hidden nodes can be expressed as:

$$h_{W,b} = f(Wx + b) \tag{4.1}$$



(a) Single-layer Auto-encoder (b) Stacked Auto-encoder

Fig. 4.3 Illustration of auto-encoder and stacked auto-encoder

where b is bias, W represents the parameters for encoder, it also can be considered as the extracted features of the input data. $f(\theta)$ is an activation function, which can be sigmoid or hyperbolic tangent *et.al.*. Then the reconstruction of the decoder can be described as :

$$y = x' = f(W^{'}h + b^{'}) \tag{4.2}$$

Usually, the W for encoder and the $W^{'}$ are constrained by $W^{T} = W^{'}$ which means the same parameters are used for encoding and decoding. However, in some cases, it is not possible to apply these constraints. The loss of the reconstruction is denoted as:

$$J(W, W^{'}, b, b^{'}; x, x') = \frac{1}{2}||x^{'} - x||^{2} \tag{4.3}$$

During the training, a back-propagation algorithm is employed to minimise the reconstruction loss. Furthermore, the power of the auto-encoder can be improved by stacking several layers together. To enhance the performance, sparse auto-encoder and denoising auto-encoder are proposed [146]. For the sparse auto-encoder, a sparsity term ρ is added to the loss function to control the activation of the feature vector to ensure the sparsity. Whereas, the notion of the denoising auto-encoder aims to add noise to the input x to confuse the encoder while

recovering the original input with the decoder in order to enable the model to extract more robust features.

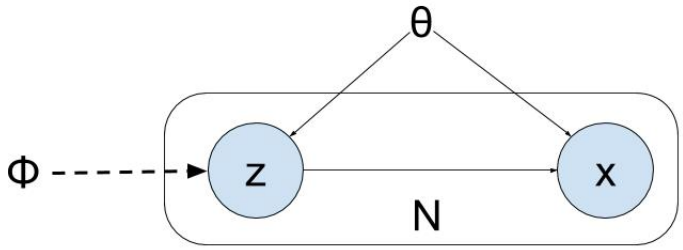


Fig. 4.4 The directed graphical model used to explain VAE in [147]

Variational auto-encoder(VAE) VAE is a generative model which introduces "variational inference" [148] into auto-encoder. As illustrated in Fig. 4.4, the objective of VAE is to train a generative model $p_\theta(x,z) = p_\theta(x|z)p_\theta(z)$, by which it can generate an x with a given latent variable z.

$$p(x) = \int p_\theta(x|z)p(z)dz \tag{4.4}$$

However, calculating the integral of the marginal likelihood is intractable since the posterior density $p_\theta(z|x) = p_\theta(x|z)p_\theta(z)/p_\theta(x)$ is intractable. Hence, an alternative probability $q_\phi(z|x)$ is proposed to replace $p_\theta(z|x)$. In the loss function, KL divergence $D_{KL}(q_\phi(z|x)||p_\theta(z|x))$ is introduced to control the similarity between $q_\phi(z|x)$ and $p_\theta(z|x)$. Thus, the log-likelihood is:

$$\log p_\theta(x) = D_{KL}(q_\phi(z|x), p_\theta(z|x)) + \mathscr{L}(\theta,\phi;x) \tag{4.5}$$

where the $\mathscr{L}(\theta,\phi;x)$ is called evidence lower bound. With the variational inference principle, the total cost of can be further transformed to:

$$J_{\phi,\theta} = D_{KL}(q_\phi(z|x)||p_\theta(z|x)) - \mathbb{E}[[q_\phi(z|x)|p_\theta(z|x)]] \tag{4.6}$$

Hence, the VAE can be considered as two part, the probabilistic encoder $q_\phi(z|x)$ and the generative probabilistic decoder $p_\theta(x|z)$. To simplify the sampling process, Gaussian distribution is usually hired for $q_\phi(z|x)$.

Convolutional auto-encoder Convolutional neural networks are proved to be very powerful in image processing as the structure is able to extract hierarchical information from the input. However, the original auto-encoder does not consider the 2D structure of the image. To adapt the auto-encoder for 2D or 3D input, the convolutional auto-encoder is proposed by [149]. Intuitively, the main change is that convolution is hired for each layer in encoder $h_{W,b} = f(x * W + b)$ and deconvolution for decoder $y = x^{'} = f(h * W^{'} + b^{'})$ instead of the multiplication in the original operation. The nature of the deconvolution is not a real deconvolution operation. In fact, it convolves the input with another set of parameters $W^{'}$. Other than that, a pooling layer is also introduced to obtain translation-invariant representations. The rest of the network, such as the reconstruction loss and training method, is consistent with other convolutional neural networks.

4.2.2 Image pre-processing and nuclei detection

Before the era of deep learning which is also known as a data-driven approach, researches invested a huge amount of efforts in solving the image-related problems with mathematical analysis. Whereas, these traditional image processing methods do not require any labels or annotations which is the obvious advantage over the deep learning methods, especially for those occasions without sufficient annotations. Several important techniques for nuclei processing is employed for nuclei detection and segmentation.

Colour deconvolution

One of the fundamental goals is the segmentation of HE stained images according to certain semantic structures. In such images, there are main colours present, pink for the cytoplasm

and blue for nuclei. However, the colours are not uniform across different slides due to the variation of concentration of the dye and other factors in staining procedures, and can even be different within the same slide due to the changes in slide thickness. As long as the two colours in the same slide are differentiable, the nuclei will be easily distinguished from the background.

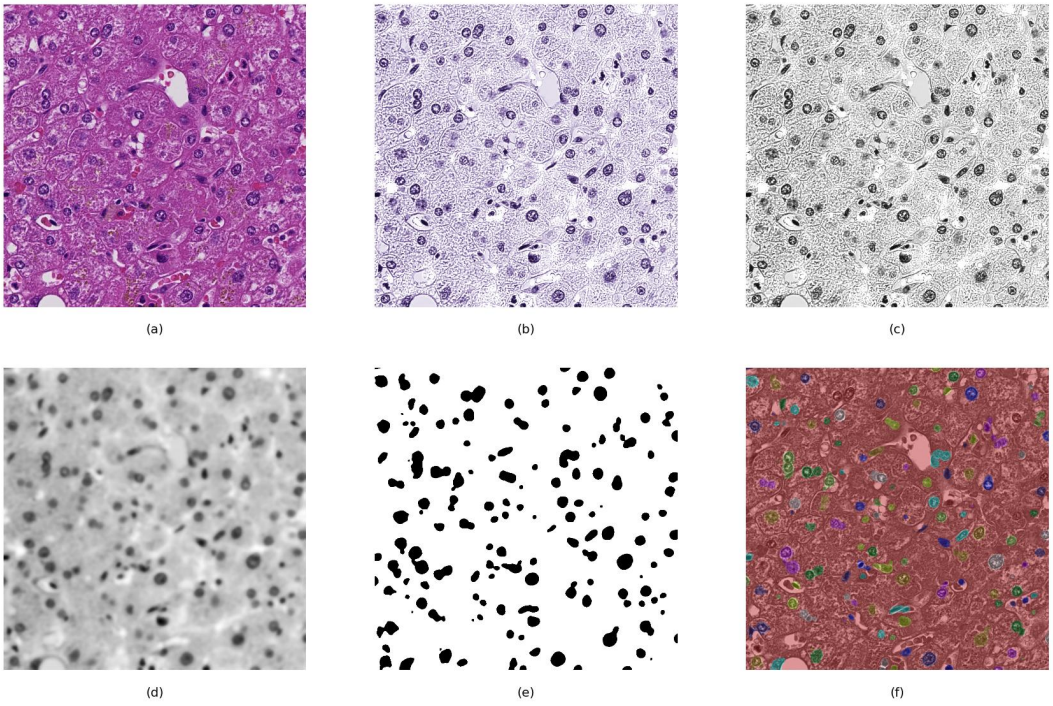


Fig. 4.5 Nuclei region proposal pipeline. (a) Original image (b) H channel image for colour deconvolution (c) Greyscale image (d) Blurred image (e) Proposed region from adaptive threshold (f) Refined results with morph-snake algorithm

In the RGB colour space, the problem can be formulated as follows: a hyperplane needs to be defined by the two main vectors representing the colour of the haematoxylin and the colour of the eosin to ensure that the projection of all the colours in the slide to this plane has the minimum reconstruction error. For an image matrix M with size $N \times N \times 3$, it can be represented as Eq. 4.7:

$$M \approx I \times B \tag{4.7}$$

where B is the bias vector with the size 2×3, and I ($N \times N \times 2$) is the intensity map of the two colours. To solve this equation, the non-negative matrix factorisation (NMF) algorithm proposed by [150] is employed. The original image will be decomposed into two images: an H channel image which concentrates on the nuclei and an E channel image which focuses on the cytoplasm. The results shown in Fig. 4.5 (b) is a separated H channel image extracted from the original image Fig. 4.5(a).

Region proposal

From the separated H channel image, a clearer view of the nuclei's distribution and shapes can be acquired. However, the noise in the image and the variation of the intensity still hinder us from segmenting all the nuclei out with one threshold. Besides, the shape of the nuclei are not always closed since the slicing operation during sample preparation will cut some nuclei into pieces. In this case, to find the centre of a nucleus, the total variation denoising (TVD) algorithm [151] is adopted on the down-sampled image to smooth the image and highlight the nuclei in the region (as shown in Fig. 4.5(d)). Afterwards, to address the intensity variation problem, a local adaptive threshold is used to separate the background from the foreground. With a basic thresholding, a possible nuclei region $I'(x,y)$ can be estimated from the image $I(i)$:

$$I'(x,y) = \begin{cases} 1, \ if \ I(x,y) > T \\ 0, \ otherwise. \end{cases}, T \in [0,1] \tag{4.8}$$

where T represents a threshold value. In comparison, the local adaptive threshold method considers the local context and adjust the threshold accordingly. T is determined by the local mean $m(x,y)$ and standard deviation $\delta(x,y)$ of a window with size $w \times w$ centred at (x,y):

$$T(x,y) = m(x,y)[1 + k(\frac{\delta(x,y)}{R} - 1)] \tag{4.9}$$

where R is the maximum of standard deviation and k is a bias in the range of [0.2, 0.5]. Fig. 4.5(e) shows an example of the result of the adaptive threshold.

Active contour for mask refining

Snake model is one of the most popular active contour models designed to detect object boundaries [152]. For the initial boundary constructed by a set of points, the energy of it can be defined as :

$$E_{Total} = E_{In} + E_{Ex} \tag{4.10}$$

E_{In} indicates the force to shrink the contour and smooth the curve C:

$$E_{In} = \int \alpha(|C^{'}|)^2 + \beta(|C^{''}|)^2 dC \tag{4.11}$$

α and β are two weighting variables that control the relative importance of the two factors. E_{Ex} represents the energy to expand the contour which can be formulated as:

$$E_{Ex} = -(|\nabla I(x,y)|)^2 \tag{4.12}$$

where $\nabla I(x,y)$ extracts the edge information of the image. Following the algorithm proposed in [153], the morphological snake model is employed to refine the mask for each patch. The advantage of the morphological snake lies in the morphological operation after each iteration which will smooth the edge produced by pre-defined operators. The initial level set comes from the binary value of the grayscale H-channel image with an adaptive threshold. The importance of the inside content and the outside content is set equal. Three times of smoothing is performed on the result of the basic active contour. However, after the process, if two contours are found in one patch, that means these two cells are too close to each other and have been mistaken as one. While, if no contour is found, that means this patch might be a background image.

The output of the proposed segmentation method is demonstrated in Fig. 4.6. The proposed region on the left is intermediate results after region proposal step, while the extracted nuclei on the right are the output after refining with the active contour algorithm.

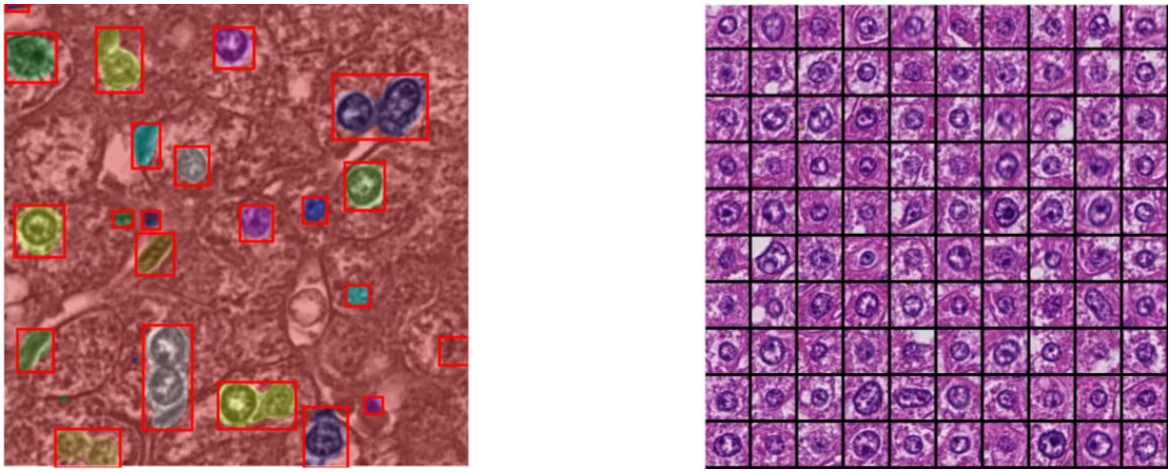


Fig. 4.6 Proposed regions (left) and extracted hepatocyte nuclei regions (right)

4.2.3 Convolutional variational auto-encoder for nuclei classification

Among the cells in the normal liver tissue, there are mainly four different types, as shown in Fig. 4.7. They are hepatocytes, kupffers cells, stellate cells and endothelial cells. About 80% of the cells are hepatocytes, and they have regular round nuclei. Endothelial cells' nuclei tend to be elongated and flattened. Kupffer cells are located in the sinusoids and have the bean or ovoid-shaped nuclei. For the stellate cells, they have spindle-shaped cell bodies and oval or elongated nuclei which are difficult to distinguish from endothelial cells.

Method

To model all the nuclei of cell types in the hepatocyte region and extract their representations, two types of unsupervised neural networks are employed, Convolutional Auto-Encoder (CAE) and Convolutional Variational Auto-Encoder (CVAE). These two networks are very similar in their structures. They both have a pair of convolutional encoder and decoder. The

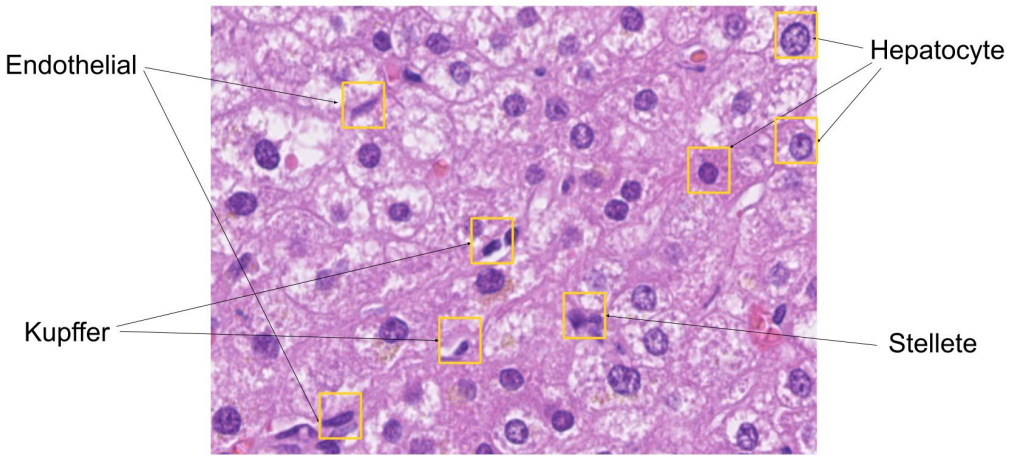


Fig. 4.7 Different types of cells in hepatocyte tissue

only difference is that CVAE has a variational inference layer instead of a linear layer before the decoder as shown in Fig. 4.8. The convolutional encoder network has no difference with other convolutional neural networks. Each layer in the network facilitates the convolutional operation to acquire the feature maps. Then the last layer of the network is a fully-connected layer which is also the input of the auto-encoder. The convolutional decoder network has the same number of layers as the encoder network, which means each layer in the encoder network has a counterpart in the decoder network. The decoder networks perform the deconvolution operation and aim to reconstruct the input image with given parameters. The loss of the network is defined as the norm of the error between the input image and the reconstructed image.

For the convolutional encoder and decoder networks, two different structures are employed to compare the performances, the Alexnet structure with pooling layers and a 6-layer fully-convolutional network (FCN) without pooling layers. The detailed parameters for FCN are listed in Table. 4.1. The loss of the networks is the binary cross-entropy of the input patch and the reconstructed patch.

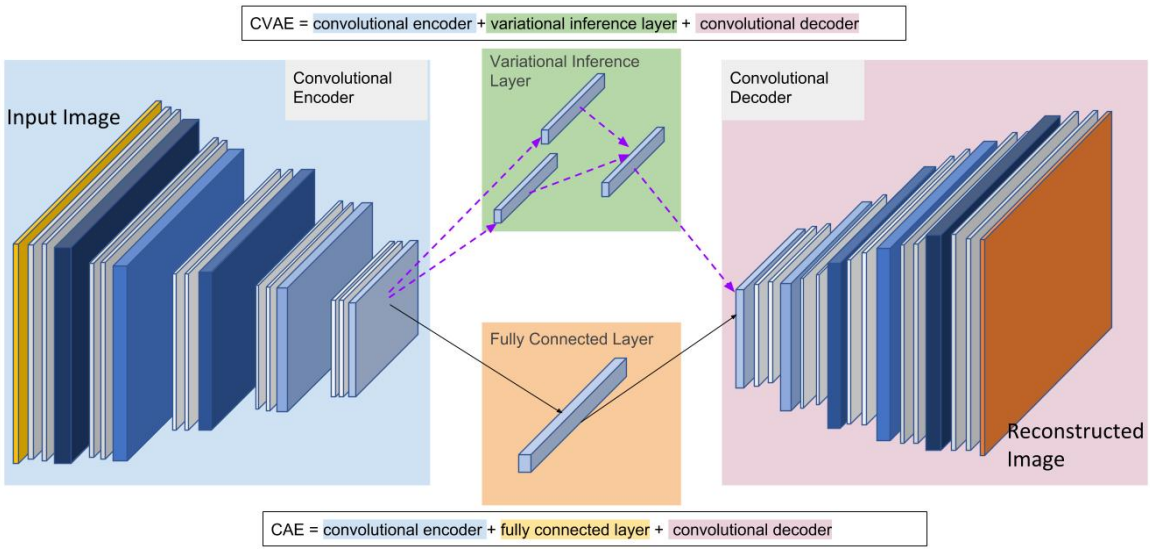


Fig. 4.8 Network structures of convolutional auto-encoder and variational auto-encoder.

Implementation details and results

The dataset for training comes from 5 different WSIs of the CRLM cases. 10 patches with the size 2000×2000 pixels are sampled from the 40X objective. After the pre-process step as described in section 4.2.2, 31833 nuclei regions are proposed. From the centre of each proposed region, a patch with a fixed size of $64 \times 64 \times 3$ is extracted and used for training the model. However, the small nuclei whose size are far smaller than the average size will be discarded, as well as those exceptionally large ones. To augment the training set, each

Table 4.1 Encoder and decoder layout of FCN

Encoder					Decoder			
Layer	Depth	Filter size	Stride		Layer	Depth	Filter size	Stride
Input	3				Input	8/16		
1. Conv	3	4×4	(2, 2)		1. Deconv	64	4×4	(2, 2)
2. Conv	16	4×4	(2, 2)		2. Deconv	32	4×4	(2, 2)
3. Conv	16	4×4	(2, 2)		3. Deconv	32	4×4	(2, 2)
4. Conv	32	4×4	(2, 2)		4. Deconv	16	4×4	(2, 2)
5. Conv	32	4×4	(2, 2)		5. Deconv	16	4×4	(2, 2)
6. Conv	64	4×4	(2, 2)		6. Deconv	3	4×4	(2, 2)
7. Conv	254	4×4	(2, 2)		7. Deconv	3	4×4	(2, 2)
8. Output	16/32				7. Output	3		

patch is flipped horizontally and vertically. At last, around 90000 patches are prepared for training. For each network structure, 8 and 16-dimensional feature vectors are extracted for classification. Fig. 4.9 demonstrates their original input images and the reconstructed images from the 16 dimensional CAE networks.

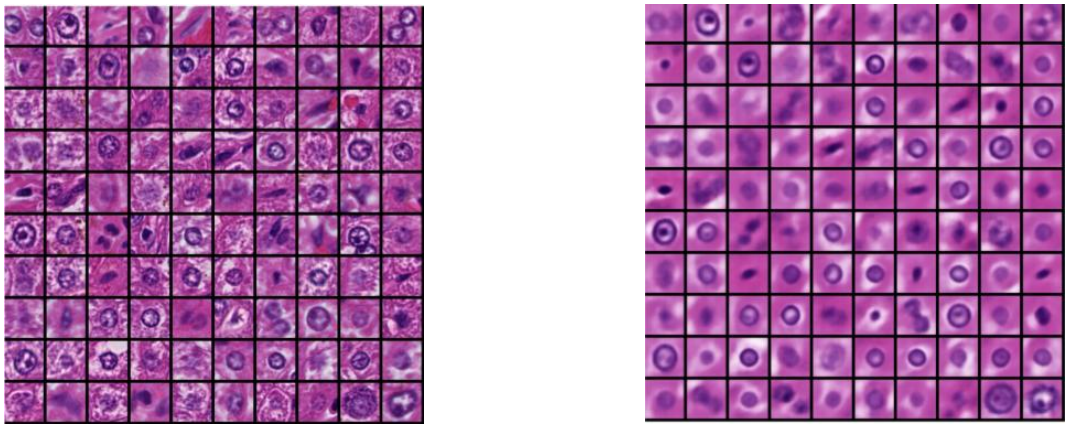


Fig. 4.9 Input image patches (left) and the reconstructed image patches by CAE (right)

The CVAE is a generative network, the network can generate a certain images by learning, with a set of pre-defined parameters. As shown in Fig. 4.10 (b-e), those nuclei are generated based on the same output vector. For each patch, only one dimension is changed comparing to the neighbour patches. Thus, similar input patches will result in similar feature vectors. With the help of the t-SNE [143] algorithm, the extracted 8-dimension and 16-dimension feature vectors are mapped to a 2D dimensional space as shown in Fig. 4.10 (a). The four colour denotes the four classes clustered by the algorithm. As it can be observed in the figure, the features of different classes clutter together, making it difficult to define the centre of each class.

From the pathologists' point of view, the proposed cells can be classified into four categories in term of their functionalities. However, according to the nuclei's visual appearance, they can be categorised into four groups as well: the hepatocytes with single nuclei are treated as an independent group; the endothelial, kupffer and stellate cells are merged as one group since they have elongate nuclei or ovoid shape; those cells with touching nuclei are labelled

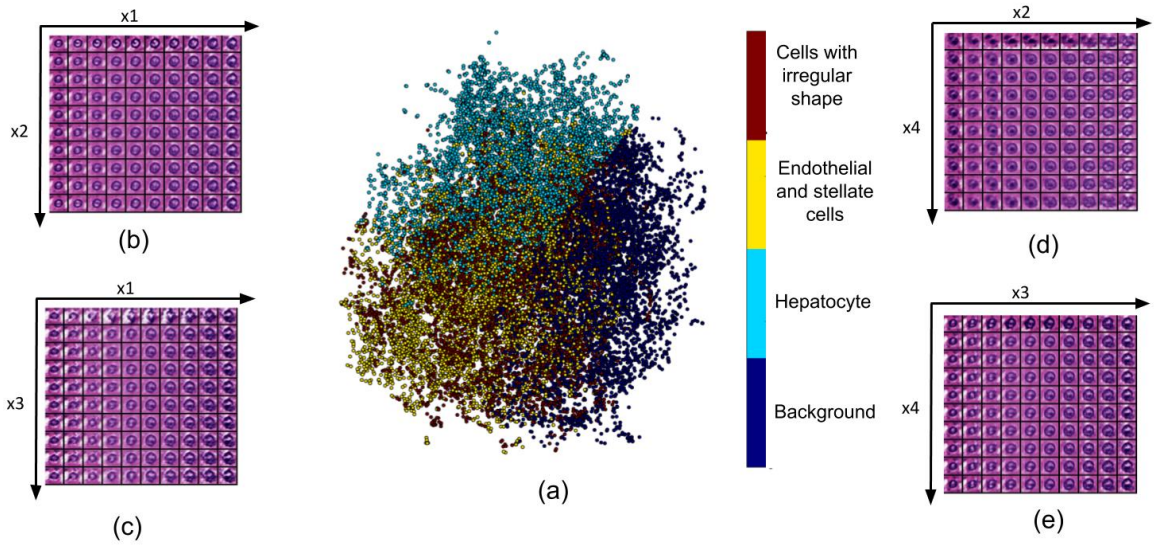


Fig. 4.10 Features of the nuclei regions on transformed space. (a): 2D visualisation of the t-SNE distribution of the feature vectors. (b)-(e): generated nuclei from 2 variants of the vector.

as another group together with other cells with irregular nuclei. Hence, these four classes are hepatocytes, endothelial and stellate cells, cells with irregular shape and background.

In this study, 16 manually picked patches from each of the class are considered to define the centre of the cluster C. The samples used to calculate the centre are demonstrated in Fig. 4.11. The labelled cells are flipped vertically and horizontally to augment the library. With the feature vectors, $\mathbf{V}$ derived from the output of the trained network, the mean vector of each class is calculated as the centroid of that class.

$$\mathbf{M} = \frac{1}{16}\sum_{n=1}^{16}\mathbf{V}_n \tag{4.13}$$

For each patch p, the Euclidean distance between the feature vector $\mathbf{V}_p$ of this patch and the centroid $\mathbf{M}_i$ of the i^{th} classes is compared as shown in Eq. 4.14 . The minimum distance will be used for the indication of the classes. The class whose centroid leads to the minimal

distance from the patch is considered as the class of the patch.

$$l = arg\min_{i}(||\mathbf{V}_p - \mathbf{M}_i||), \quad i \in \{1,2,3,4\} \tag{4.14}$$

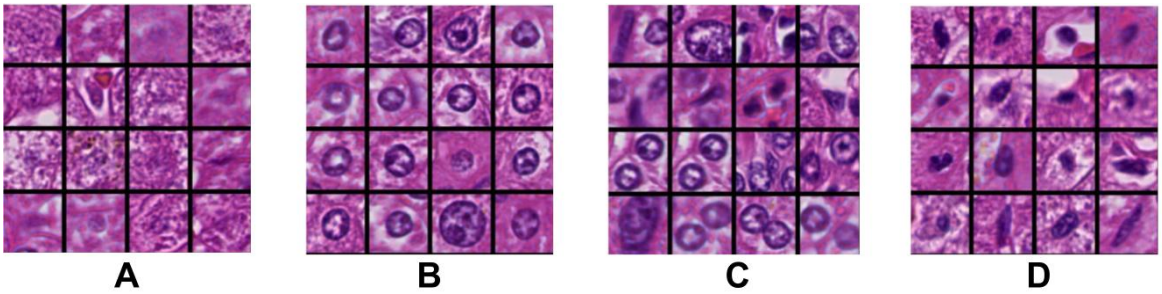


Fig. 4.11 Samples that used to extract the centre of the features. A: Background B: Hepatocyte C: Cell with irregular shapes D: Endothelial and stellate cells

Evaluation and discussion

To evaluate the performance of the different network structures and dimensions of the feature vectors, 3000 patches are randomly sampled from the dataset and annotated. Precision and recall are calculated for each structure, as reported in Table 4.2, from which three observations can be made. FCN structure outperforms the AlextNet structure on both accuracy and recall. The increase of the feature vectors' dimension will improve the performance. The CVAE structure can output better results than the CAE structure.

Table 4.2 Evaluation of the classification precision and recall

	8 dimension		16 dimension	
	Precision	Recall	Precision	Recall
CAE-Alexnet	82.4%	75.2%	83.9%	74.3%
CAE-FCN	83.1%	74.9%	83.8%	76.1%
CVAE-Alexnet	85.4%	78.8%	86.2%	80.2%
CVAE-FCN	86.6%	79.1%	87.9%	80.5%

The study presented in this section demonstrates a novel workflow for nuclei classification with insufficient annotation data. Efforts have been made to exploit the possibility of tackling

the classification problem by combining the weakly supervised method and traditional image processing algorithm. Even though the accuracy of the proposed weakly supervised method is far from satisfying, the proposed method can play a critical role in generating annotations, which can significantly reduce the manual workload.

4.3 US-net for robust nuclei instance segmentation

Comparing to natural objects, the detection and segmentation of nuclei seem to be much easier due to their simple structures and homogeneous properties in representation. However, despite the fact that the topic of nuclei detection and segmentation has been researched for decades, there is still no publicly available trained model that support universal nuclei detection across HE slides produced in different labs and conditions. Before the wide adoption of deep neural networks, conventional nuclei detection methods often use the statistical or geographical features of images to generate the seeds.

In this section, a robust model is proposed for nuclei detection and segmentation. This model, referred to as US-net, benefits from a concise, yet efficient architecture, which consists of a nuclei detection network and a segmentation network. It involves a workflow that dynamically integrates the regression output of nuclei location and the end-to-end output of the semantic segmentation to enhance the performance of both networks. The main contribution of this section is two-fold: i) a novel and robust deep neural network architecture for instance segmentation of the nuclei in HE stained histopathology image; ii) an enhanced focal loss that can help deal with the class imbalance and accelerate the training is designed.

4.3.1 Background and preliminaries

U-net [154] is a baseline network for many segmentation methods even though it is originally proposed for biomedical image related tasks. The basic structure for U-net and its derivative

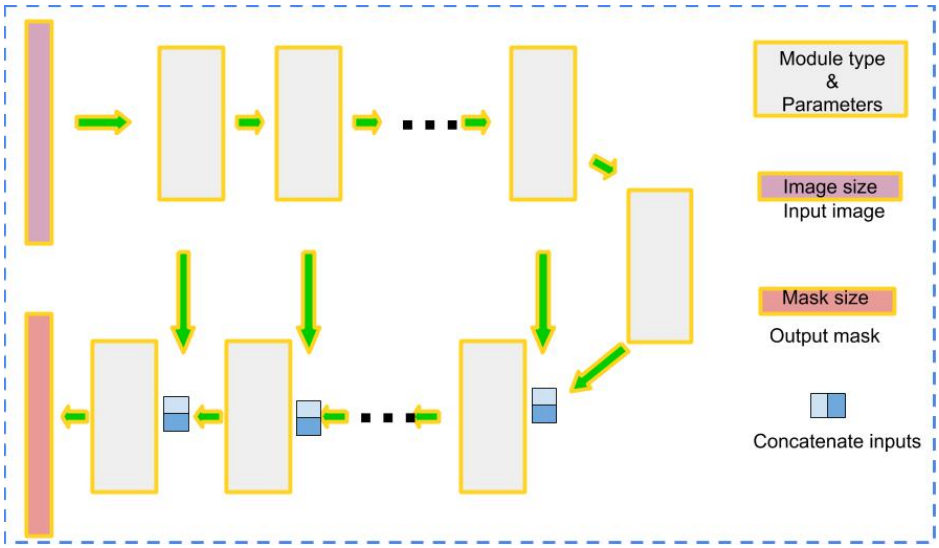


Fig. 4.12 Basic framework of the U-net

networks is demonstrated in Fig. 4.12. The key idea of U-net is to pass the low-level features directly to high-level so that during the up-sampled/deconvolved process, the low-level features can contribute to the final decision making. In other words, the importance of low-level features is emphasised, which explains why it is prevalent among biomedical image processing. For the original U-net, the blocks are filled with convolutional modules. If these blocks are replaced with Residual module and Dense module, they are Res-Unet and Tiramisu-net [155] accordingly.

Single shot detection (SSD) [156] is a powerful one-step object detection network comparing to the two-step networks such as R-CNN [157] and faster R-CNN [158]. With one shot, the SSD model is capable of locating the object and classifying the object as well. Fig. 4.13 illustrate the network structure, the multibox detector [159], anchor box and the data structure of SSD network. In SSD, it employs a multi-box detector to replace the region proposal operation in the two-step object detection networks, which can be explained with Fig. 4.13 (C & D), for each point on the feature map, a set of anchor boxes will be employed for the location prediction. For each anchor box, the location information will be optimised

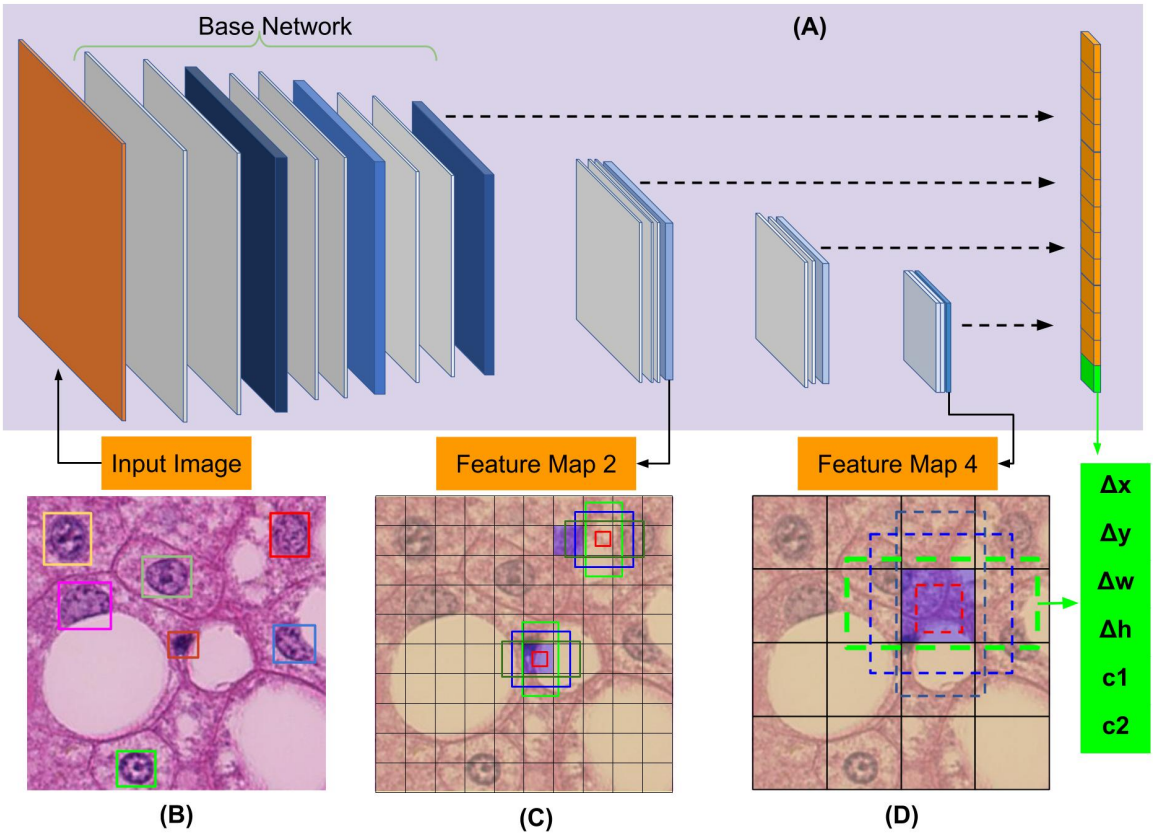


Fig. 4.13 Network structure of SSD network

in a regression manner, while, for the category information, it will be calculated with softmax. In the end, all of the predictions will be concatenated for predicting the final result.

4.3.2 US-net architecture

US-net is a specialised network that is designed especially for nuclei detection and segmentation. As shown in Fig. 4.14, the structure of US-net is very compact, composed of a segmentation and a detection branch, sharing the same backbone network. In practice, the proposed network takes advantage of the powerful end-to-end semantic segmentation ability of U-net [154] structure and the excellent object detection and classification performance of SSD [156], precisely RetinaNet [160], to achieve an instance segmentation results with the help of a post-processing sub-network for refinement.

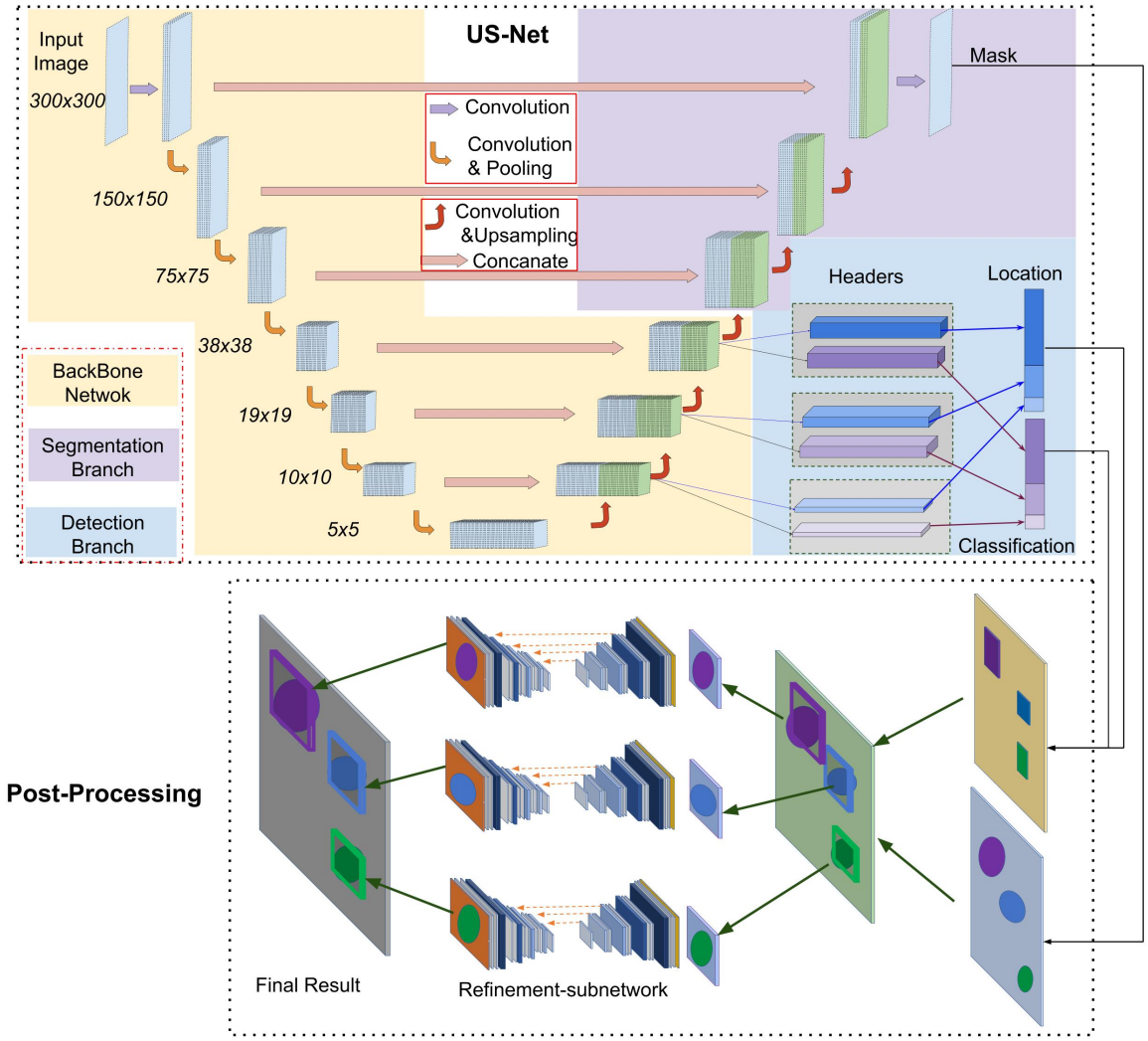


Fig. 4.14 Network structure of proposed US-net with post-processing sub-networks

The overall objective of the network extends the loss of MultiBox objective [159]: given an input image I and its corresponding segmentation masks s , location information l and its class information c , the loss L can be decided by the following function:

$$L(I,c,l,s) = L_{conf}(I,c) + \alpha L_{loc}(I,l) + \beta L_{seg}(I,s) \tag{4.15}$$

where the parameters α and β control the relative importance of the loss components.

The term $L_{seg}(I,s)$ which is defined with the $L2$ norm, and it helps to achieve the segmentation results.

$$L_{seg}(I,s) = avg(||s-\hat{s}||^2) \tag{4.16}$$

where $\hat{s}$ is the segmentation results and s is the ground truth information. The loss L_{conf} measures the confidence scores of the binary class (nuclei or not) of the detected boxes with an adapted version of Focal loss [160] :

$$L_{conf}(I,c) = \eta_{\hat{s}c}(1-\hat{c})^{\gamma} log(\hat{c}) \tag{4.17}$$

where $c \in \{0,1\}$ denotes the ground truth information for classification, $\gamma(\geqslant 0)$ is the focusing parameter and $\eta_{\hat{s}c}$ is associated with the output of the segmentation results:

$$\eta_{\hat{s}c} = \sqrt{\frac{S_{all}}{\hat{S}_c}} \tag{4.18}$$

where S_{all} represents the number of the pixels in ground truth s, while $\hat{S}_c$ denotes the number of pixels that equals to c in output of the $\hat{s}$. The term $L_{loc}(I,l)$ in Eq. 4.19 calculate the location regression loss of the multi-boxes given the ground truth information l with a Smooth $L1$ loss as defined in [159].

$$L_{loc}(I,l) = \frac{1}{N}\sum_{i=1}^{N}\sum_{p\in(x,y,w,h)} smooth_{L1}(l_i^p - \hat{l}_i^p) \tag{4.19}$$

For the i^{th} bounding box location information l_i, (x,y) is the center of the box while h and w represent the height and width of the box. With the Soomth $L1$ loss defined as:

$$smooth_{L1} = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & |x| \geq 1 \end{cases} \tag{4.20}$$

Since the output of US-net can only achieve a relatively rough segmentation result, thus, post-processing steps are necessary to realise an instance level segmentation of the input image. The post-processing part is also described in Fig. 4.14. With the rough location and rough segmentation result, another network with U-net structure that consisting of 8 convolutional layers is built for further refining the detected regions. The function of the post-processing network is similar to the attention mechanism, which is to refine the results from a local scope.

4.3.3 Experiments and evaluation

DATASET

The training dataset employed in the experiments come from the Segmentation of Nuclei in Images Contest (SNIC)[161] and the MICCAI MoNuSeg [162, 163] . There are 32 patches with size 600×600 pixels from SNIC and 30 patches with size 1000×1000 pixels from MoNuSeg. Both of them have instance level annotation masks. The proposed model works with input patches of size 300× 300. Hence, the images from the original dataset are cropped to size 400×400 with a fixed step size of 200. After pre-processing, 878 patches are acquired in which 650 patches are used for training and 228 patches are reserved for evaluation. For all the nuclei from different organs, they are treated as the same which means no category information will be attached to each nucleus since the accurate detection and segmentation of the nuclei is the main focus here. In addition to the 300×300 patches, another dataset with patch size 48×48 is created for the refinement sub-network in the post-processing stage. These patches are cropped from the scaled bounding boxes area and then resized to 48×48.

Implementation Details

In the U-net part, there are six down-sampling layers and six up-sampling layers connected by a bottleneck layer. The block size for all the layers is 4. For the purpose of performance

comparison, different blocks such as Resnet blocks and Densenet blocks are introduced to replace the convolutional blocks. In the SSD part, the detectable objects' (nuclei) size is constrained to the range of 20∼128 by using the feature maps from the last 3 layers of the base network. The corresponding feature map sizes are $38 \times 38, 19 \times 19, 10 \times 10$. Two anchor box aspect ratios (1×1 and 1×0.75), and two scales (0.8 and 1.2) are considered in the experiments which make 4 different anchor boxes for each point in the feature maps. That makes up the 7620 default anchor boxes for an input image. In the training phase, the parameter α is set to 1 while β is set to 0.1. Two different optimisers are employed for two different branches. For the SSD branch, Adam optimiser [164] with a learning rate of 0.001 is employed, while for the U-net branch, the optimizer is SGD with a learning rate 0.0001, momentum 0.9 and weight decay 0.0001.

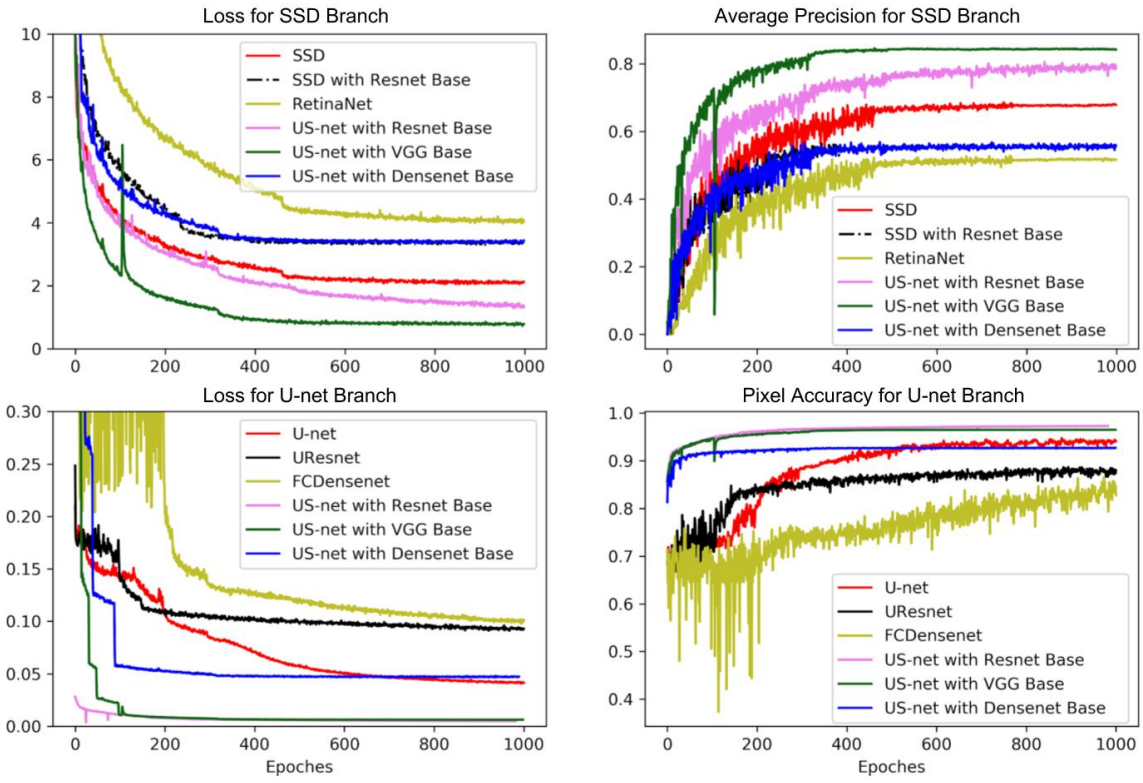


Fig. 4.15 Losses and evaluation accuracies for different network structures

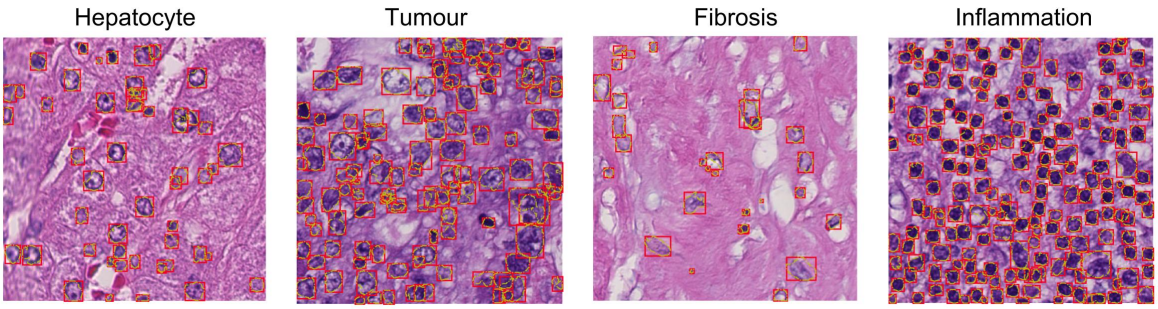


Fig. 4.16 Results of the proposed method on the CRLM dataset

4.3.4 Results and discussion

The evaluation of the results of the proposed networks takes two aspects into account, the evaluation of semantic segmentation and the evaluation of detection results. For the segmentation part, PA is calculated in terms of the number of accurately predicted pixels out of the total number of pixels. The evaluation metric for object detection part is the same as the interpolated mAP used for VOC dataset [165], which has been introduced in Sec 2.4. In Fig. 4.15, losses for different branches of the US-net as well as the AP/PA for evaluation are demonstrated along the training process. It can be observed that the US-net performs significantly better than either of the branches individually. Besides, it should be noted that the inclusion of the Densenet blocks or Resnet blocks into the framework does not necessarily increase its performance due to the low complexity of the features in the image. Furthermore, the trained model is applied to the histopathology images from CRLM cases used in the previous section. The results presented in Fig. 4.16 demonstrates the outstanding performance of the trained network which verifies the decent robustness and transferability of the proposed framework.

4.4 Multi-scale fusion for semantic segmentation in colorectal liver metastasis

The accurate segmentation of the tumour tissue regions in histopathological images is of crucial importance for cancer pathology analysis. However, due to high morphological variance in histopathology images, especially along the border regions where cancer cells interfere with normal tissue, it is challenging even for pathologists to identify a clear boundary between tumours and normal tissue, let alone an automated tumour segmentation system. In this section, an innovative framework is presented to segment tumour areas of CRLM at the cell level in a semantic manner, by integrating the features from the spatial and statistical information of cells. With the rough annotations from pathologists, deep neural network architectures, including a cell-level model and a tissue-level model, are trained to classify patches from the whole slide scan images. Afterwards, based on the trained models, a growing-style algorithm is designed to segment the images by closely investigating the statistical and spatial properties of the cells. In a comprehensive evaluation against the ground truth created by professionals, the proposed framework demonstrates significant improvement over the conventional deep convolutional neural networks operating at the tissue level alone. Technically, the contributions of this work are to integrate the multi-scale results to achieve a more accurate pixel-level semantic segmentation results with the proposed tissue region growing algorithm.

4.4.1 Framework and proposed method

An overview of the proposed framework is illustrated in Fig. 4.17. This framework integrates two workflows from the cell level and tissue level in order to adequately exploit the most representative and distinctive morphological features on the two levels in a complementary manner.

At the tissue level, the PatchCNN model proposed in Chapter 3 is employed to make a rough prediction of tissue region segmentation. The advantage of starting with a tissue level prediction is that a larger reception field of the whole patch can be covered so that the model can include more contextual information to assist the prediction. Like other works on tissue level segmentation, the output boundary of this tissue level models is coarse and the details in the boundary, which may be of clinical relevance, need to be better demonstrated. Here, the integration with the cell level workflow is considered to enable a pixel level segmentation.

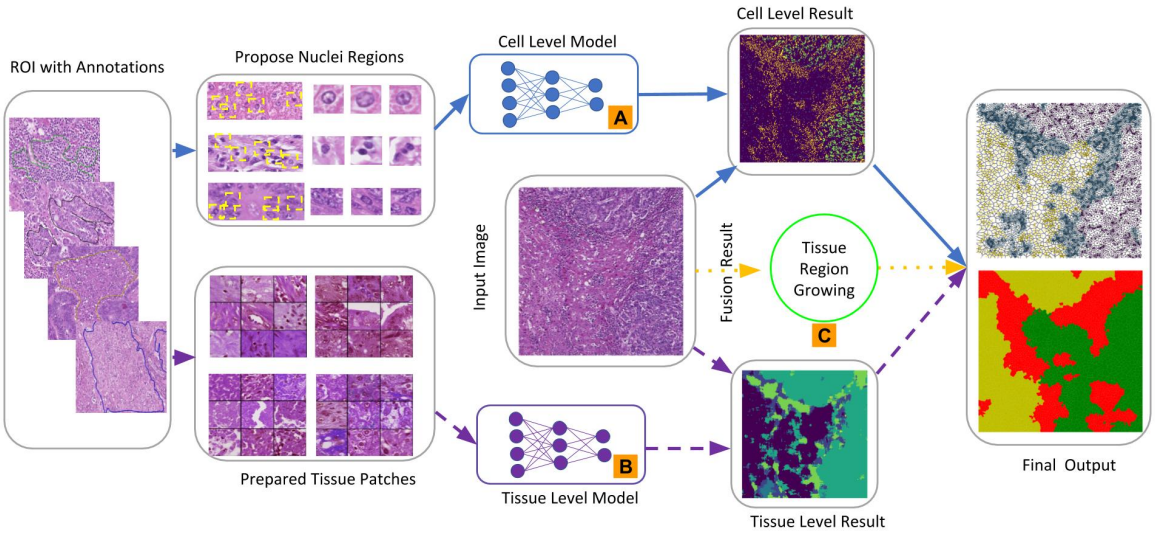


Fig. 4.17 An overview of the proposed fusion framework

At the cell level, the nuclei segmentation and classification model (US-net) proposed in Section 4.3 are employed. The purpose of this model is to classify the potential nuclei regions. Although the proposed network achieved excellent results, without context information, the result still contains a significant amount of biased data which will introduce noise to the results. Here, the lack of inclusion of contextual information is complemented by the tissue level workflow. Meanwhile, the cell level model is more sensitive to the changes in the border, which can lead to a more detailed prediction for segmentation. The integrated workflow of the framework is summarised as below:

Step 1. Process raw images and associated annotations, and prepare training data for the models.

Step 2. Train the cell level model and tissue level model separately (shown as blocks A and B in Fig. 4.17).

Step 3. Apply the trained **tissue level model** on an input image.

Step 4. Apply the **cell level model** on the image patches for a cell level prediction.

Step 5. Resize the tissue level prediction to fit with the input image's size, if the prediction by the cell level model agrees with that of the tissue level model, then mark these nuclei with corresponding labels. Mark the rest of the nuclei on which an agreement cannot be achieved as uncertain.

Step 6. Run the **Tissue Region Growing algorithm** (Block C in Fig. 4.17) to relabel those uncertain nuclei.

Step 7. Remove outliers and refine the segmentation results.

Tissue region growing for the integration of tissue and cell level segmentation After the pre-process and nuclei detection, the distribution of the detected nuclei can be converted to a graph using Delauney tree [166] as shown in Fig. 4.19. Based on the results from cell level and tissue level models, each node will be assigned a temporary label to indicate if the results from two levels agree or not. Here, the tissue region growing algorithm is designed to determine the membership of those nodes with uncertain labels.

Consider a simple example as illustrated in Fig. 4.18, the red dots belong to set $\Re$, representing nuclei with uncertain labels. The others (black and green) belong to set Φ including dots whose the labels are agreed by the two models. Θ is a subset of Φ which includes the nearest K nuclei to a nucleus r with a certain label. For each undetermined nucleus, four properties are calculated: the size of the nucleus S^r, the average size of the

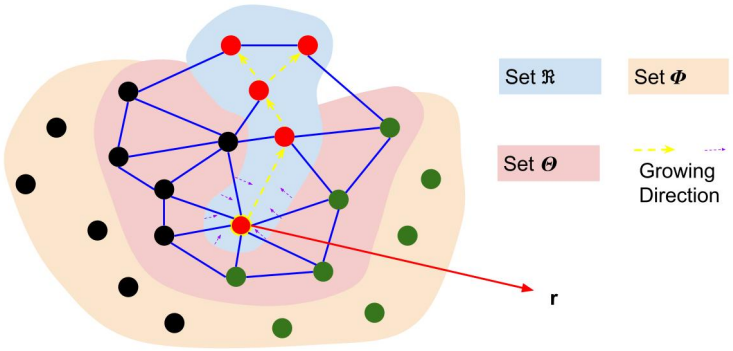


Fig. 4.18 An illustration of the proposed Tissue Region Growing algorithm

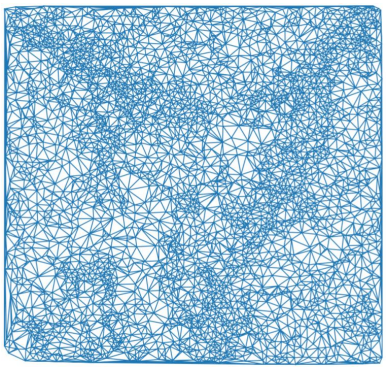


Fig. 4.19 Delauney tree of the nuclei's location

nearest K^r nuclei regions aS^r, the average distance to the nearest K nuclei's region centers D^r, and the average of K nearest nuclei regions' $D^{r,i}$, $(i = \{1,2,3...K\})$ value aD^r. For each class of nuclei with label C_i, the average value of these four properties are denoted as S_i^Θ, aS_i^Θ, D_i^Θ, aD_i^Θ, where i indicate the index of the label in set C. For the nuclei in the set Φ, the four properties are denoted as S_i^Φ, aS_i^Φ, D_i^Φ, aD_i^Φ. The distance of unlabelled nuclei r to the class C_i is defined as W_d in Eq. 4.21.

$$W_d(r,C_i^H) = \sqrt{\sum_{G\in(S,aS,D,aD)} \frac{(G^r - G_i^H)^2}{G_i^{H\,2}}}, H \in \{\Phi,\Theta\} \tag{4.21}$$

With the joint assessment of the four statistical properties in W_d, the tissue region growing algorithm assigns class memberships to the uncertain regions iteratively based on their nearest neighbours in a growing style. The iterative algorithm is described in Algorithm 1. The criteria for classification includes two parts. For those regions close to only one kind of nuclei, the goal is to decide if they can be treated as the same kind according to the distance to the statistical properties of this category. If they are not similar enough, the growing will stop. A threshold value λ controls the similarity of the region and its neighbours. To make sure of the convergence of the algorithm, a growing rate of μ is introduced. If the growing stops, the threshold value will be adjusted to reactivate the growing. For those nuclei whose

Algorithm 1 Tissue Region Growing

Precondition: Initial threshold λ, growing rate μ, regions with certain label Φ, regions with uncertain label $\Re$

repeat
  repeat
    for each $r \in \Re$ **do**
      Locate nearest K certain nuclei to r as set Θ
      if Labels $\hat{C}$ of Θ contains multiple classes **then**
        Find the $\hat{C}_m$ which leads to the minimum value in set Θ $W_d(r, \hat{C}_m^{\Theta})$
        Update r's label with $\hat{C}_m^{\Theta}$
      else if Label $\hat{C}$ only contain one class $\hat{C}_n$ in Θ **then**
        Find the one with the maximum value $W_d(\hat{r},\ \hat{C}_n^{\Phi})$ in set Φ
        if $W_d(r,\ \hat{C}_n^{\Phi}) < \lambda W_d(\hat{r},\ \hat{C}_n^{\Phi})$ **then**
          Update r's label with $\hat{C}_n$
        end if
      end if
    end for
  until Φ does not change
  Increase the value of λ by μ
until $\Re = \emptyset$

neighbours include multiple kinds of nuclei, the goal is to attribute the region to the right one. The criteria are the distances of the region to each category's statistical centre.

4.4.2 Dataset preparation

In this section, several different datasets are employed for this research. The dataset for the tissue level model is the same as those used in Chapter 3 where tissue level model is described and trained. While, the situation for the cell level are kind of complex. The model for nuclei detection and segmentation are trained on the public dataset as described in the last section. However, for the fusion framework, a cell level CRLM dataset are prepared. This dataset is made from 120 homogeneous tissue patches manual selected from the tissue level dataset on the $40X$ level with the patch size of 448×448 pixels. Each of them has an overall label that indicates its class membership. These patches are used to extract the nuclei for

constructing a CRLM classification dataset, in which each nucleus image is of size 128×128 and is cropped from the nucleus centre.

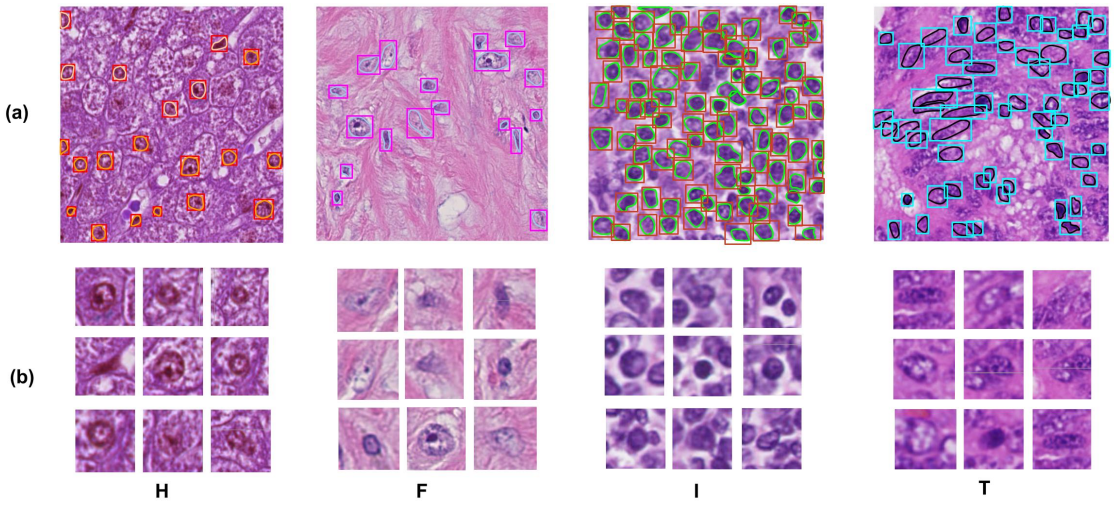


Fig. 4.20 Examples of the detected nuclei on the large patch (a) and the extracted nuclei regions (b)

From 30 patches, over 5000 nuclei with high confidential scores are extracted. Among them, 3500 are used for training and 1500 are used for validation. In the last, a 7-layer fully convolutional network is established for the classification task. The evaluation on the independent testing dataset indicated that 96.28% of the nuclei (over 1500 samples) can be accurately classified. Fig. 4.20 illustrate the examples of the patches and the extracted cell level dataset from them. In the last, to test the proposed segmentation methods, the same testing dataset containing 50 ROIs as employed in Chapter 3 are used for evaluation.

4.4.3 Implementation details

The fusion stage take place after the tissue level results and cell level results is acquired separately. The main process of the fusion stage is the implementation of the tissue region growing algorithm. The Fig. 4.21 demonstrates the intermediate results of the fusion process. The parameters λ and μ is set to 0.1 and 0.2 respectively. To evaluate the proposed framework, the same evaluation metrics are considered: PA and MIoU.

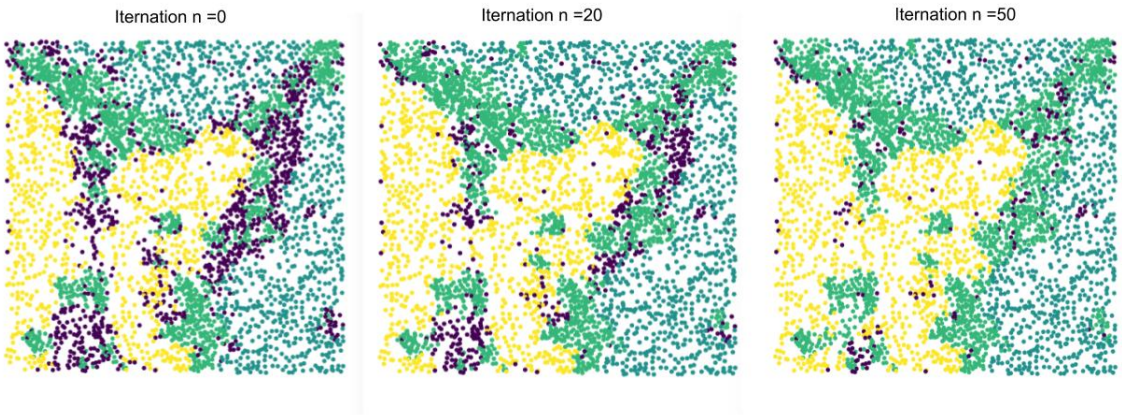


Fig. 4.21 The outcomes of the intermediate steps of the tissue region growing algorithm. The purple points indicate the nuclei with the uncertain label, the yellow, green, blue ones represent H,F/I, T tissue types respectively.

4.4.4 Results and discussion

The quantitative analysis on the testing set with the metrics are listed in Table 4.3. It is reported in the table that comparing to the tissue level models, the proposed method with merged results from the two levels can improve numerical value of both the PA and MIoU. Furthermore, from the distribution of the nuclei in Fig. 4.21, it is easy to found that most of the purple points (nuclei with uncertain labels) fall in the border area of two types of tissue where the tissue region growing algorithm works.That's explained why the results improved.

Table 4.3 Quantitative segmentation result (PA & MIoU) of proposed methods

	PatchCNN	Resnet-34 + CRF	PatchCNN + Superpixel	Proposed fusion method
PA	78.12%	80.47%	80.87%	81.83%
MIoU	58.16%	59.33 %	59.74%	61.07%

In addition, a few example of the segmentation results are demonstrated in Fig. 4.22. Except for the segmentation result, the nuclei map of the input image are demonstrated as well. The results in the last row of Fig. 4.22 expose the bad performance of the cell level model. With carefully inspection, it can be found that the nuclei in this kind of fibrosis have confusing appearance because they are very similar to some of the tumorous nuclei. Besides,

there are some tumorous nuclei mixed inside as well. Meanwhile, the tissue level result could not provide enough information for the fusion process, which lead to the overall bad performance. In summary, the cell level model can help to enhance the segmentation result to some extent, but the framework still has some constraints to overcome.

4.5 Summary

In this chapter, the overall objective is to make more accurate segmentation the tumour border region with a weakly supervised method by utilising the information from the cell level. To this end, my research was carried in three-folds. Firstly, by leveraging the image processing method, we modelled the nuclei in the hepatocyte area through a generative model, variation auto-encoder. Then, a novel network is proposed to do the instance-level nuclei detection and segmentation In the last, a multi-scale fusion framework is proposed to increase the segmentation accuracy and resolution. Extensive experiments have been carried out and the results confirmed that the improvement of the proposed method.

Fig. 4.22 Examples of segmentation results. A : original input image; B: ground truth mask; C: Resnet-34 + CRF; D: PatchCNN + Superpixel; E: Cell level result F: Proposed fusion method.

Chapter 5

Weakly Supervised Models for Pixel Level Histopathology Image Analysis

Pixels are the basic units of the digitalised image on a computer, and all operations on these images will eventually lead to numerical changes in the pixels. However, the concept of "pixel level" in this thesis does not refer to the specific pixels in the histopathology images, since the information contained in one pixel are closely associated with the magnification level, for example, one pixel in an image from $5\times$ magnification level contains information from a $8px \times 8px$ patch at $40\times$ magnification level. Hence, the "pixel level" in this study refers to the details of the pixels at the highest magnification level, which is the $40\times$ magnification level. Even though it might not be essential for clinical diagnosis with pixel level segmentation, it could provide a more accurate calculation on the WSIs by assigning pixels with different classes. In this chapter, several different weakly supervised methods are proposed to address the challenge of insufficient annotations at the pixel level. Firstly, the augment training methods are investigated, and two different approaches for generating matched training pairs are proposed to solve the insufficient pixel level annotations of the training data. Secondly, virtual staining is explored to transform the HE stained images to an IHC stained image in a

weakly supervised manner. The high contrast characteristic of the IHC style staining can be facilitated for use with an auxiliary visualisation method.

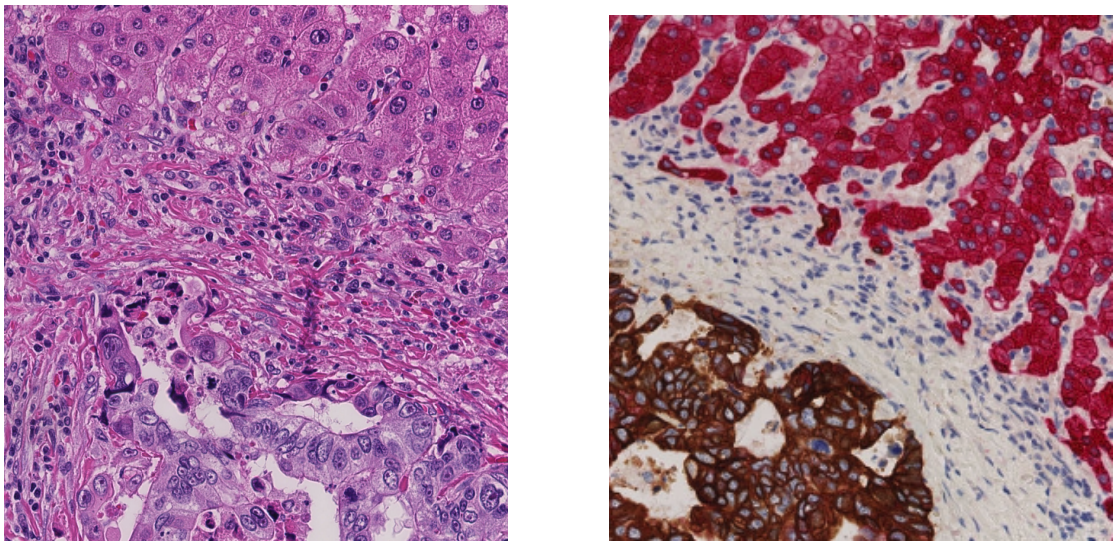
5.1 Overview

Pixel level segmentation aims at training an end-to-end model that can transform the input image to another image with different pixel values thereby indicating different types of tissue. In this context, the pixel level segmentation has similar purposes as virtual staining. Virtual staining, by its literal meaning, uses computerised algorithms to create the artificial effect of staining without physically tampering with the slide. This means virtually converting the staining of the image to another kind of staining through computer algorithms according to the semantic features of the structures. In fact, virtual staining is a post-processing step that can generate other stained versions of the same slide based on the digital scan of a real slide stained using a traditional dyeing method. The aim is to provide different staining effects that can highlight different relevant histological and cell features on the same slide. From this point of view, the objective of pixel level segmentation and virtual staining is almost identical. Hence, in this study, virtual staining is employed as an auxiliary method to assist the pixel level segmentation.

5.1.1 Problem statement

In clinical practice, there are other methods, mainly IHC staining, that a pathologist can use to highlight the cancerous or other cell and tissue components of interest. As shown in Fig. 5.1, IHC staining allows the visualisation of specific proteins on a tissue slide using targeted antibodies and subsequently detecting the bound antibodies using chromogens of different colours. In such a way, by using protein-binded antibodies that are selectively expressed in the cancer cells, these become readily distinguishable from other tissues using high-contrast

chromogen staining, as shown in Fig. 5.1 (b). Despite its superior features compared to HE, there are obstacles that preclude the wide application of IHC staining in clinical practice - the high cost of antibodies, autostainer machine equipment, and the complex laboratory process. Unfortunately, these issues are currently unresolvable. In most cases, pathologists have to rely on HE or other relatively cheaper and more generally available staining methods for their diagnosis and limit the use of IHC to a very small number of selected slides. For this reason, the majority of computer models and systems for histopathological cancer grading are designed for segmenting and classifying the tumour tissue based on HE imaging.



(a) HE stained image (b) IHC stained images

Fig. 5.1 Matched HE and IHC stained images patches from the same region of two adjacent WSIs

Fig. 5.2 illustrates how the IHC stained image can be facilitated in different tasks. As can be observed, the segmentation results can be easily acquired from IHC stained images with thresholds. In this case, if the HE stained images can be converted to IHC stained images, the problem can be addressed easily. From the computer vision aspect, the transformation is asymmetric due to the information contained in these two kinds of images not being equal. With all these channels being decomposed, it can be observed that it is relatively easy to compose HE stained images through a linear combination of the channels from IHC images.

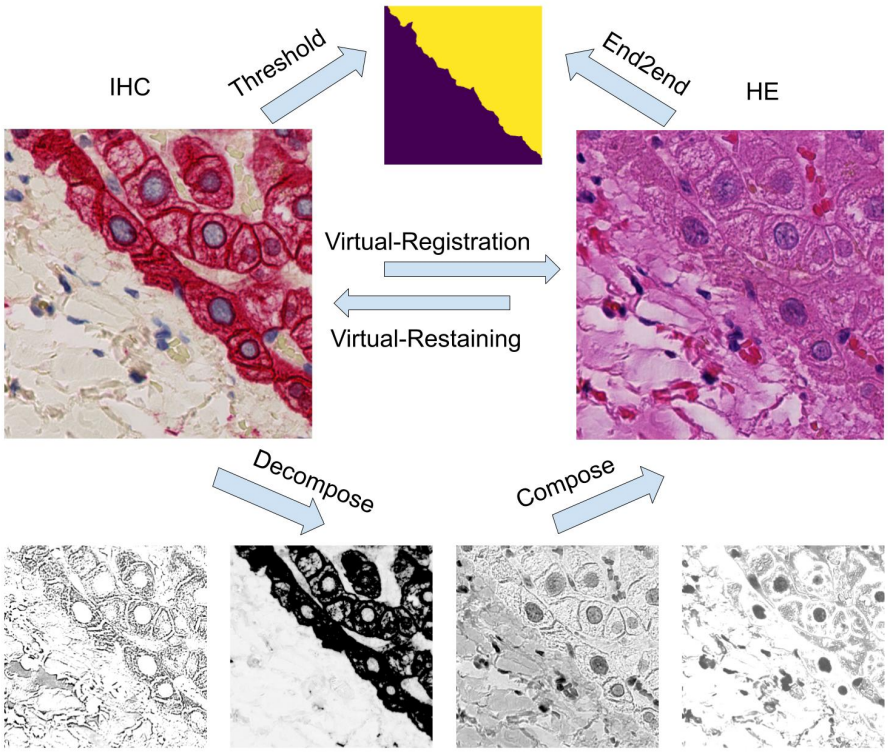


Fig. 5.2 Pixel level problem statement and task overview

But it would be rather difficult to compose a IHC image with linear combinations of these channels since they do not contain the same amount of information. Hence, the transformation from IHC to HE is referred to as virtual registration, while the reverse transformation is called virtual restaining. The term "virtual restaining" is used in order to emphasize the difference between another kind of "virtual staining" which transforms the unstained images to stained images.

The transformation between a IHC and HE stained image can be considered as a special kind of style transfer problem [167] that only transfers the colour coding to the original images while keeping the structural information consistent. It is also similar to the image colourisation problem which will re-colourise the image with new colours [108]. However, this method needs to be based on the excellent performance of semantic segmentation. Based on the tightly related common features of images with different stainings, the most relevant problem is the unpaired image-to-image translation that will transfer the original image

using the style of the reference image. Therefore, if there is a way to model the pairwise relationship between the tissue morphology in HE and IHC stained conditions, it is then possible to predict one from the other. Recently, with the continuous progress of deep learning with different image processing tasks, it has become possible to 'translate' one image to another with a different style. Such technology brings the opportunity to produce an image virtually stained with one dye from another, for example, producing IHC images from images dyed with relatively cheaper and widely available staining techniques such as HE. This may be very useful in many clinical diagnostic and artificial intelligence (AI) based applications, such as improving the effectiveness of pathological examination by reducing the eyeballing time in the visual screening of the slides, and increasing the segmentation and classification performance of AI models. I.e. if the HE slides can be virtually IHC stained, with simple post-processing, the generated IHC images can provide highly precise segmentation of the tissue regions of interest.

What is more interesting is that given an original image acquired with one stain, virtually restained images with one or more other dyes can now be generated from exactly the same sample. Currently, only one or two chromogens can be employed on the same sample for brightfield microscopy, which is the standard in histopathology practice. However, it is often the case that several IHC stainings need to be performed on the same tissue region to extract multiple features of diagnostic relevance. In such cases, pathologists perform serial cuts from the same tissue block, each between 3µm and 5µm thick, and stain them with different techniques and antibodies. This introduces inevitable inter-slide variability in the cell and tissue structures, as while performing serial cuts along the z-axis, neighbouring tissue regions are typically similar but do not exactly match each other, preventing cell-level segmentation and colocation analyses across slides. Consequently, the differently stained sections do not represent exactly the same tissue sample. With the help of virtual staining, multiplexing with two or more multiple stains on the same sample becomes possible.

5.1.2 Challenges and objective

The large size of unannotated histology image data has posed critical challenges for its understanding and analysis. Motivated by the great success of deep learning models applied in different tasks in natural image analysis, more and more deep learning algorithms and systems are being designed for histopathology image analysis. Dozens of fruitful outcomes have been achieved based on deep convolutional neural networks (DCNN), especially using patch-based methods. With a sufficient number of annotated patches, a DCNN will be trained to detect cancerous tissue patches against other benign tissue components. However, there are intrinsic limitations in the current methods. A major issue is the high dependence networks on annotated training sets that are often inadequate both in terms quantity, due to the expensive manpower consumption required for the job, and in terms of quality, because of inter- and intra-observer subjectivity. Moreover, when multiple tissue types are present in a region intertwined together, it becomes impossible even for experienced pathologists to create accurate annotation masks, such as the HE stained image region shown in Fig. 5.1 (a). For this reason, the lack of adequate and accurate training data is the most challenging problem.

Another challenge that the pixel level segmentation task faces is the difficulty in evaluation. As stated above, it is almost impossible to make an accurate annotation on the tissue border manually. In this case, the evaluation of the transformed image or segmented image will be difficult. Besides, for most of the annotations conducted by experts, they are more focused on the semantic meaning. While, the masks from the IHC stained images will maintain the small objects as well, which makes the segmentation results noisy and coarse.

5.1.3 Preliminaries

Generative Adversarial Networks [93], as an important branch of deep learning, is gaining more and more attention from the researchers around the world. A significant amount of

investigations have been made to explore the potential of GAN for natural images related tasks like image synthesising, image super-resolution, and style transfer [107, 108]. However, modelling the patterns in the histology images is a particularly complicated task for generative models since tissue morphology can be treated as orderless texture patterns. Although there are underlying regulations that control the growth pattern of the cells, yet the rules are limited and unknown to human experts. In this case, the generative adversarial networks are often served as a data augmentation method that helps generate more tissue regions. This is one of the most intuitive application of generative models in histology image related application. But the potential of applying generative model is not limited to that, more innovative and promising researches are on the way. In the following, the basic GAN model and its derivatives (or important applications) are introduced first.

Generative Adversarial Networks

The fundamental idea of a GAN, as shown in Fig. 5.3, is to train the generative model G_{θ_g} that can generate fake images from the noise z though learning the real images, to fool the discriminator D_{θ_d} which can discriminate real/fake images. Once the discriminator $D\theta_d$ cannot tell the fake ones from the real, it means that the network has learned to model the distribution of the input data appropriately. The mathematical description of overall loss is:

$$\mathscr{L}_{gan}(G,D,x,z) = \mathbb{E}_{x\sim p_{data}}[\log D_{\theta_d}(x)] - \mathbb{E}_{z\sim p_z}[\log D_{\theta_d}(G_{\theta_g}(z))] \tag{5.1}$$

where θ_g and θ_d are the parameters of the generator network and the discriminator network. The objective of the networks is to maximise the discrimination accuracy and minimise the error of generated images so that the generated data can have identical distribution as the real data.

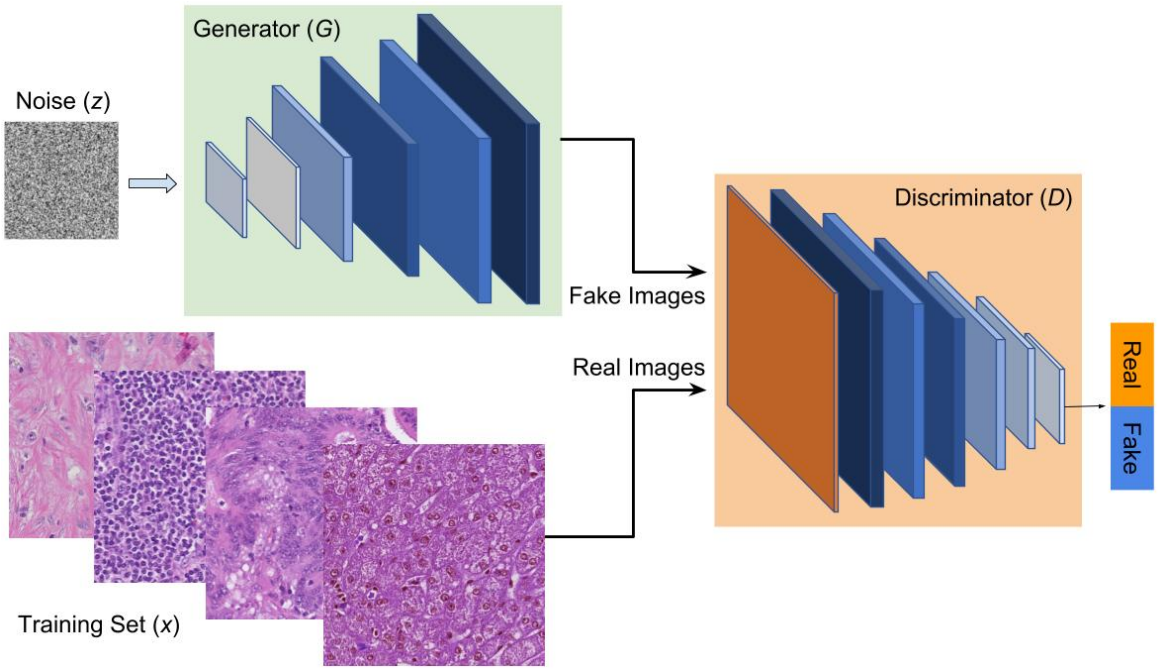


Fig. 5.3 Network structure for generative adversarial network

GANs for paired image translation

Based on the original GAN networks, conditional GAN[168] has been proposed to take into account certain constraints that can help to improve the truthfulness of images, in addition to the imaginary information. Instead of generating an image from imaginary, conditional GAN will apply some constrictions/ conditions on the model and further guide the model on how the images should be generated. The conditions encode the labels of the generated images which contain category information. Furthermore, pix2pix GAN, as a paired image-to-image translation method, provides segmentation masks as conditions to generate images on a pixel level [169]. From the image translation point of view, the pix2pix model translates the images from abstract representation to real images while keeping the semantic meaning.

GANs for unpaired image translation

Among the above mentioned paired image-to-image translational models, an obvious drawback is the requirement for corresponding masks on the input images. Thus, a few un-

paired image translation models are proposed including SimGAN [97], CoGAN [170] and CycleGAN[171]. Although the images are unpaired and from different domains, they shared similar semantic and structural features, which provides the foundation for unpaired image translation. For example, in CycleGAN, the model tried to translate a horse to a zebra and an apple to an orange. The objects before and after the translation have to demonstrate similar semantic structures to make the translation meaningful.

5.2 Augmented training for pixel level segmentation

Following the discussing above, if the problem of paired training data is tackled, it will be not that challenging to train a pixel level end-to-end model. Therefore, in this section, two augmented training methods are proposed as weakly supervised methods to generate training pairs for the end-to-end model in order to achieve a pixel level segmentation result.

5.2.1 Image synthesis for data augmentation

It is acknowledged that the tissue image is a special kind of texture image [172]. Therefore, inspired by the work in texture segmentation [173], the tissue image synthesis approach is employed to generate the dataset for pixel level model. The workflow of the image synthesis is explained in Fig. 5.4. The input images are the tissue patches from different types of tissue, but all the patches have uniform visual features within the same patch. By assigning them to different segments of the mask, the new image will be synthesised. However, to obtain a natural appearance, all the patches are normalised to smooth the border between different types of tissues.

Regarding the mask, the main challenge is to make the border have a natural appearance. Therefore, the masks acquired from Chapter 3 with superpixel algorithm is utilised to generate masks for the syntheses process. Then, different kinds of tissues are filled in the

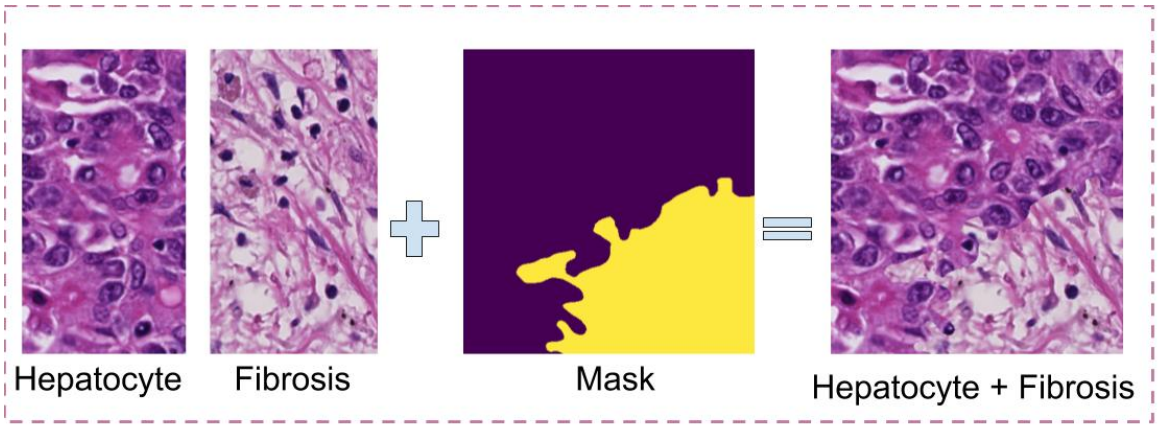


Fig. 5.4 Image synthesis framework for data augmentation

mask generated by superpixel to synthesis new patches. In order to ensure the contextual information and avoid the over-segment problem, the masks will have no more than 3 regions and the proportion of each region should account more than 1/3 of the whole area. However, such constraints would introduce new problems as well. For example, the proposed method could not synthesis those patches containing small tumours inside, because these small tumours are discrete objects which are difficult to synthesis. In this context, image transformation based methods are presented to synthesis more realistic image pairs.

5.2.2 Image transformation for data augmentation

The high-contrast feature of IHC reveals that there are explicit borders between different kinds of tissues. This inspires us to seek the possibility of segmenting the tissue border in pixel level. As discussed in Fig. 5.2, if the IHC stained images can be transformed to HE stained images, there will be enough paired images for training. Therefore, in this section, the virtual registration task is addressed through a novel GAN-based framework as demonstrated in Fig. 5.5. Considering of importance of the overall contextual information over the low-level textural details and the constraints of the computation ability, the proposed framework is divided into 2 sub-networks, the virtual registration (VR) block (B) and the conditional virtual magnification (CVM) block (A). The VR task is conducted on 20$\times$ magnification

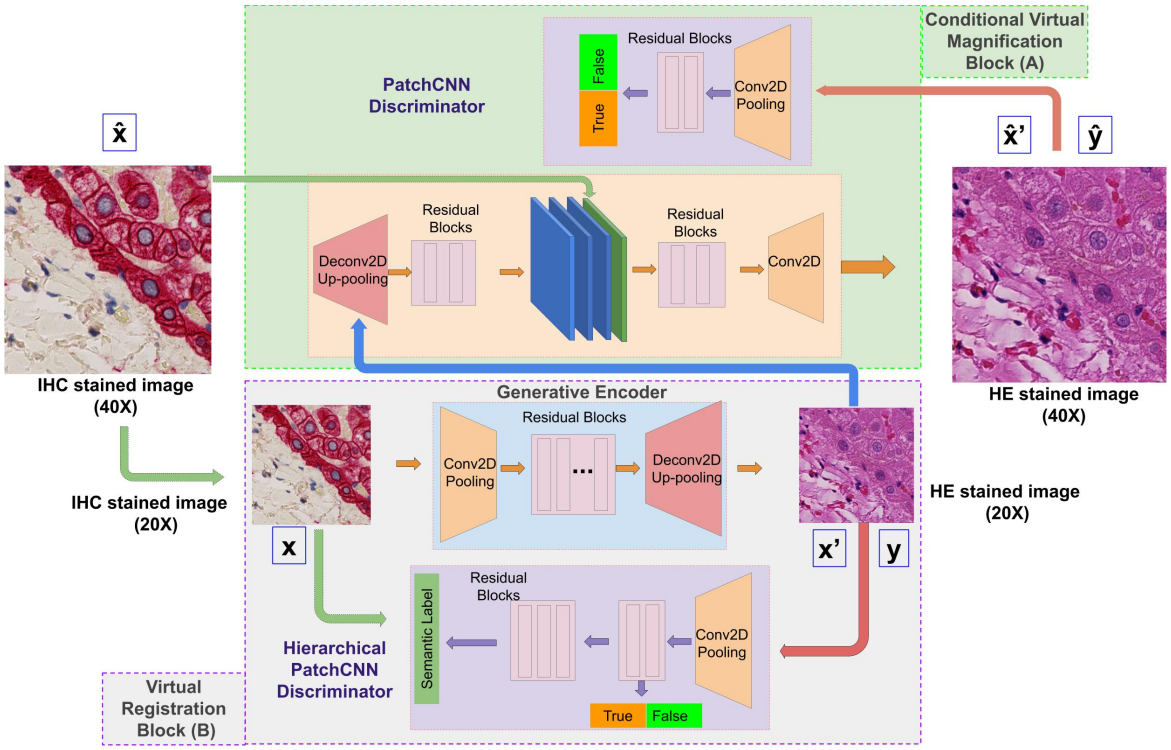


Fig. 5.5 Augmented training framework for pixel level segmentation. $\hat{x}$ and x denote the input IHC stained images in 40× and 20 × magnification level accordingly. Likewise, $\hat{y}$ and y for HE stained images. x' is output of virtual registration module while $\hat{(x)}'$ is output for virtual magnification model.

level concentrating on transforming the IHC stained images to HE stained images, while, the CVM task will help to restore the resolution on 40× by using the details from the original 40× IHC image.

Virtual Registration

The VR block is composed of two modules: the generative encoder G_{vr} inside takes in IHC stained images and converts them into HE stained images; the hierarchical discriminator D_{vr} inside is in charges of improving the quality of generated HE images with regards to the semantic label c and the training samples of the HE stained images. To ensure the quality of the generated image, a hierarchical discriminative loss is proposed. The loss contains two parts: the low-level true/false PatchCNN loss regarding how real the patches are and the

high-level semantic classification loss indicating the semantic category of the whole patch. The definition of the hierarchical loss is:

$$\begin{aligned}\mathscr{L}_{hier}(G_{vr},\, D_{vr},\, x,\, y,\, c) = \mathbb{E}_{y\sim p(y|c)}[logD_{vr}(y)] + \mathbb{E}_{y\sim p(y)}[logD_{vr}(y)] \\ +\mathbb{E}_{x\sim p(x|c)}[log(D_{vr}(G_{vr}(x)))] - \mathbb{E}_{x\sim p(x)}[log(D_{vr}(G_{vr}(x)))]\end{aligned} \tag{5.2}$$

In addition to the hierarchical discriminative loss, identity loss and photorealism loss are hired in the module as well. The identity loss proposed by Taigmen et al. [174] is introduced to preserve the colour composition especially when the input image is very close to the output image domain [171]. The identity loss can be obtained by feeding the HE stained images to the input. The definition of identity loss is demonstrated as follow:

$$\mathscr{L}_{idt}(G_{vr},\, y) = \mathbb{E}_{y\sim p(y)}[||(G_{vr}(y) - y||_1] \tag{5.3}$$

Furthermore, to keep the structural information unchanged, the photorealism loss [167] is introduced to to ensure the texture structure details of the input image unchanged. Photorealism loss uses Matting Laplacian transform to measure the structural differences and is defined as :

$$Pho(p,\, q) = \sum_{k=1}^{3} q_k^T [M_p] q_k + \sum_{k=1}^{3} p_k^T [M_q] p_k \tag{5.4}$$

where k denotes the RGB channels while M_p/M_q is the transformed matrix regarding to the input images pair p/q. Hence, the photorealism loss for VR block can be formulated as following:

$$\mathscr{L}_{pho}(G_{vr},\, x) = Pho(G_{vr}(x),\, x) \tag{5.5}$$

The overall loss of the VR block is:

$$\begin{aligned}\mathscr{L}_{vr}(G_{vr},\, D_{vr},\, x,\, y,\, c) = \mathscr{L}_{hier}(G_{vr},\, D_{vr},\, x,\, ,y,\, c) + \\ \tau\mathscr{L}_{idt}(G_{vr,\, y}) + \omega\mathscr{L}_{pho}(G_{vr},\, x)\end{aligned} \tag{5.6}$$

Where τ and ω are the weight parameters that regulate the importance of relative losses. Thus, the overall objective is to find the optimal generator $\hat{G}_{vr}$ and discriminator $\hat{D}_{vr}$:

$$\hat{G}_{vr},\ \hat{D}_{vr} = \arg\min_{G_{vr}} \max_{D_{vr}} \mathscr{L}(G_{vr},\ D_{vr},\ x,\ y,\ c) \tag{5.7}$$

Conditional Virtual Magnification

The main task of the CVM block is to boost the resolution of the transformed HE stained images according to the local textual details of the corresponding 40× IHC image, which is to virtually magnify the 20× HE stained images to 40×. The structure of the CVM block are presented in Fig. 5.5 (A), which is very similar to the VR block. It has a conditional generator G_{cvm} and a discriminator D_{cvm} as well. During the training, four losses are hired to regulate the up-sampled image to ensure both the local textual details and the semantic meaning of the patch. They are discraminative loss $\mathscr{L}_{gan}$ on the final output image $\hat{x}'$, photorealism loss $\mathscr{L}_{pho}$ between the input IHC image $\hat{x}$ and $\hat{x}'$, identity loss $\mathscr{L}_{idt}$ for y and $\hat{y}$ and the $\mathscr{L}_{L1}$ loss between the input image x' and the resized output $\bar{y}$.

$$\begin{aligned}\mathscr{L}_{cvm}(G_{cvm},\ D_{cvm},\ x',\ \hat{x},\ \hat{y}) = \mathscr{L}_{gan}(G_{cvm},\ D_{cvm},\ x') + \rho\mathscr{L}_{pho}(G_{cvm},\ \hat{x}) + \\ \eta\mathscr{L}_{idt}(G_{cvm},\ \hat{y}) + \varepsilon\mathscr{L}_{L1}(G_{cvm},\bar{y})\end{aligned} \tag{5.8}$$

where ρ, η, ε are three weight parameters, the $\mathscr{L}_{gan}(G_{cvm},\ D_{cvm})$ is defined as :

$$\mathscr{L}_{gan}(G_{cvm},\ D_{cvm},\hat{x},x',\hat{y}) = \mathbb{E}_{\hat{y}\sim p(\hat{y})}[logD_{cvm}(\hat{y})] - \mathbb{E}_{x'\sim p(x')}[log(D_{cvm}(G_{cvm}(x',\hat{x})))] \tag{5.9}$$

Similarly, the overall objective is to find the optimal generator G_{cvm} and discriminator D_{cvm} :

$$\hat{G}_{cvm},\hat{D}_{cvm} = \arg\min_{G_{cvm}} \max_{D_{cvm}} \mathscr{L}(G_{cvm},\ D_{cvm},\ x',\ \hat{x},\ \hat{y}) \tag{5.10}$$

5.2.3 Dataset collection and preparation

Two different datasets are employed for different augmentation methods. The image synthesis approach is purely base on the HE stained images, while the image transformation methods will need additional information from IHC stained images. Hence, the two different datasets will be introduced separately.

Dataset for image synthesis

Patches with homogeneous patterns are cropped from the HE stained WSIs with a patch size $1024px \times 1024px$ are collected. Only 4 classes are considered in the end, they are H, F/I, T and N. However, the F and I will be treated as different classes during the synthesis and training. Furthermore, to make the synthesised patches possess a more natural appearance, normalisation is applied on the homogeneous patches before sticking them together. But during the training, the whole image will be augmented to simulate the real situation. In total, there are 3000 masks are cropped, 2000 HE stained image patches (400 for each class) are collected. During the training phase, both the synthesised images and the images with homogeneous appearance are used as input. The proportion of them is set to equal.

Dataset for image transformation

Matched HE and IHC stained slides (2-plex CK19/CK18) were scanned at $40\times$ with Hamamatsu NanoZoomer slide scanner. On HE stain, all cell types show varying tones of dark blue and pink. CK19 stains in brown colorectal cancer cells using DAB chromogen, while CK18 stains in red the surrounding and benign liver cells (hepatocytes) with AP chromogen. Immune cells, stromal cells and extra-cellular matrix (fibrosis) are negative for both cytokeratine (CK) stainings. Immunohistochemical sections are also counterstained with haematoxylin (H) to visualise the background context of cells and tissue components on the slide.

As histological sections are 4 μm thick, there is always some degree of inter-slide variation between the matched slides. The differences may not be apparent at low magnification levels, for example, $10\times$, but become evident at higher magnification levels, e.g. $40\times$. In this study, the dataset are cropped from $40\times$ magnification level with a patch size of $1024px \times 1024px$ pixels. However, for the transformation block, the patches are from $20\times$ magnification level, while for segmentation block, the input images are on $40\times$. Given the overall objective of this study and the main colours in the IHC stained slides, the main classes considered in this section are *Hepatocyte (H), Fibrosis/Inflammation (F/I), Necrosis (N)* and the tissue border regions which contain more than one kind of tissue, *Tumour & Fibrosis (TF), Hepatocyte & Fibrosis (HF), Tumour & Necrosis (TN), Tumour & Hepatocyte & Fibrosis (THF), Tumour & Fibrosis & Necrosis (TFN), .*

The total number for both kinds of staining is 2400 with 300 for each type. The masks for the IHC images are derived through an image processing framework proposed in Sec. 4.2. After separating different channels with colour deconvolution algorithms, thresholds are applied on the channels to get the seeds area. Morphological operations are performed to filter out those small objects and small holes to keep the integrity of the semantic mask. In the last, all the pairs are examined by the experts to improve the quality of the data. To ensure the generalisability of the transformation for IHC images to HE images, both of the IHC images and the HE images are augmented to simulate the staining variance of the slides from different labs. This is one of the critical steps for handling staining variances across different labs. Meanwhile, the HE images for the input of the discriminator are normalised to reduce the dimension of the target domain. After these augmentation steps, the trained network will obtain the ability to normalise the output image as a side product. It is especially true during the identity loop when the same image is distorted and resized as input but has a corresponding normalised output. More importantly, it will greatly reduce the efforts when composing different small tiles into a large whole slide image. Other image augmentation

techniques (mainly geometric transform) are also employed to enlarge the number of the sample for training a more robust model.

5.2.4 Implement details and results

The implementation details of the proposed methods are detailed here, which reports the network structures and the parameters for the weights.

U-net for end-to-end segmentation

The ultimate goal of this task is to train an end-to-end model that is capable of segmenting the input patch at a pixel level. Hence, after solving the insufficient data issue, the task remaining is about the end to end model. The U-net structure introduced in Sec 4.3 is employed here again because the model needs to take into consideration of both the low-level edge information at the local context and the high-level semantic meaning of the overall patch. To ensure a larger receptive field of the model, the U-net constructed here is much deeper than the model employed in Sec. 4.3.

Network structures

The two blocks in the framework are implemented separately instead of integrating into one. The VR block is implemented first. The encoder network in transformation has 6 layers of Resnet blocks except for the convolutional header and deconvolutional footer. The hierarchical structure is used to fuse the information from a different level of the network. The true/false label will give feedback upon the low-level textual details after 2 Resnet blocks with $8\times$ down-sample rate while the high-level semantic label will work after another 2 Resnet blocks. In the CVM block, the idea of adversarial training is also employed as well. However, the discriminator is simple PatchCNN which is composed of two Resnet blocks and two convolutional layers. In contrast, the conditional super-resolution network is more

complex. As described in Fig. 5.5 (B), the IHC stained images are provided as a condition after a deconvolutional block and 2 Resnet blocks. The up-sampled features will keep the same dimension as the IHC images so that they can be concatenated. Afterwards, all these channels will be further fused through the 2 Resnet blocks and the convolutional block. For the U-net, constrained by the computational capability and the desire for a deeper network, the final network is constructed by 8 convolutional layers with 32 $\times$ down-sample rate.

Parameters and optimiser

There are several weight parameters in the proposed transformation based framework. All these parameters pose critical challenges for network optimisation. In this task, they are determined by trails of different combinations. The optimal setting for the τ and ω in VR block are 1, 5 and 2, the ρ, η and ε in CVM block are 2, 5, 1. Regarding to the optimiser, both the VR block and the CVM blocks employ Adam since they share the same adversarial structure. Whereas, the U-net segmentation network use SGD for the optimisation with learning rate 0.0001 and dropout rate 0.3.

Post-processing

The post-processing step is a very important step for both the mask preparation and output refinement. In the transformation based method, the masks acquired from the image processing method often contains a large amount of artefacts and noises. Likewise, the outputs of the U-net require further refinement as well. Hence, in this study, the same approach and parameters are facilitated for all the post-processing steps to ensure a fair comparison.

5.2.5 Results and discussion

The results of the data augmentation procedures are demonstrated in Fig. 5.6 and Fig. 5.7, corresponding to the synthesised and the transformed image pairs. As can be observed, in the

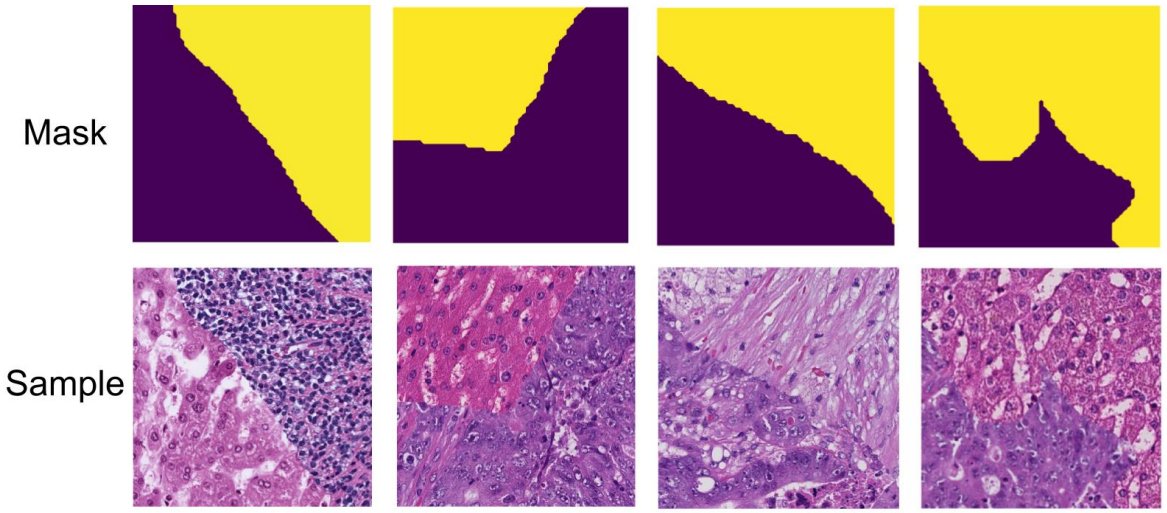


Fig. 5.6 Example of synthesised image for augment training

synthesis based method, the masks mainly address the borders for large homogeneous areas. On the contrary, the masks in transform based method have more sophisticated appearance. However, when comparing the synthesised or transformed images with the real HE stained images, the differences between them are more obvious. The synthesised images contain no semantic information in the patch and introduced artificial border to the image. While, the transformed images have no apparent difference with the real HE stained images except for the colour coding: the virtual stained images tend to be uniform across the whole patch.

To keep the integrity throughout the whole research, the evaluation dataset and metric are kept the same across different chapters. The numerical evaluation results on the testing dataset are demonstrated in Table 5.1 along with the results from previous chapters. At the same time, the corresponding qualitative results are plotted in Fig. 5.8 for visual assessment.

Table 5.1 Comparison of segmentation results (PA & MIoU) among different models

	PatchCNN	Fused	U-net1	U-net2
PA	80.87%	81.83%	81.32%	84.71%
MIoU	59.74%	61.07%	61.72%	63.02%

PatchCNN: PatchCNN with Superpixel as post-processing from Chapter 3.5 Fused: Fusion model from Chapter 4.3 U-net1: U-net with the training dataset from image synthesis U-net2: U-net with the training dataset from image transform

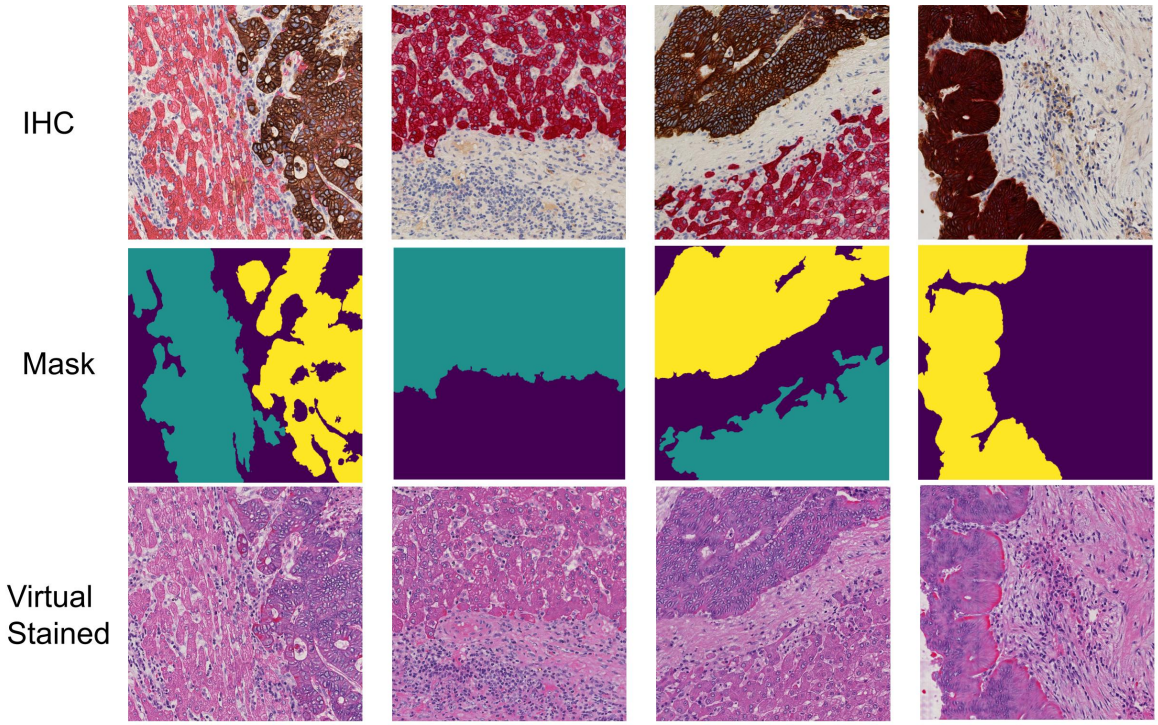


Fig. 5.7 Examples of transformed image for augment training

The numerical results in the table suggest that both of the proposed augment training methods will help to boost the accuracy of the segmentation results. However, the U-net model trained with the transform based training pairs greatly outperform the model trained with synthesis based pairs and the methods proposed in previous chapters. The reason for the unsatisfying results with the synthesised model can be attributed to the quality of the images. For the other models, the main reason of the improvement is the providence of extra information. The results in Fig. 5.8 reveal more reasons about the differences in the performance. For the patches with obvious borders, the performances of different methods are almost the same. Whereas, for the patches containing small tumours or those regions with complicated morphology, the performances vary a lot. Hence, it can be concluded that the proposed transform based weakly supervised method will efficiently help to supplement the information for a better pixel level segmentation.

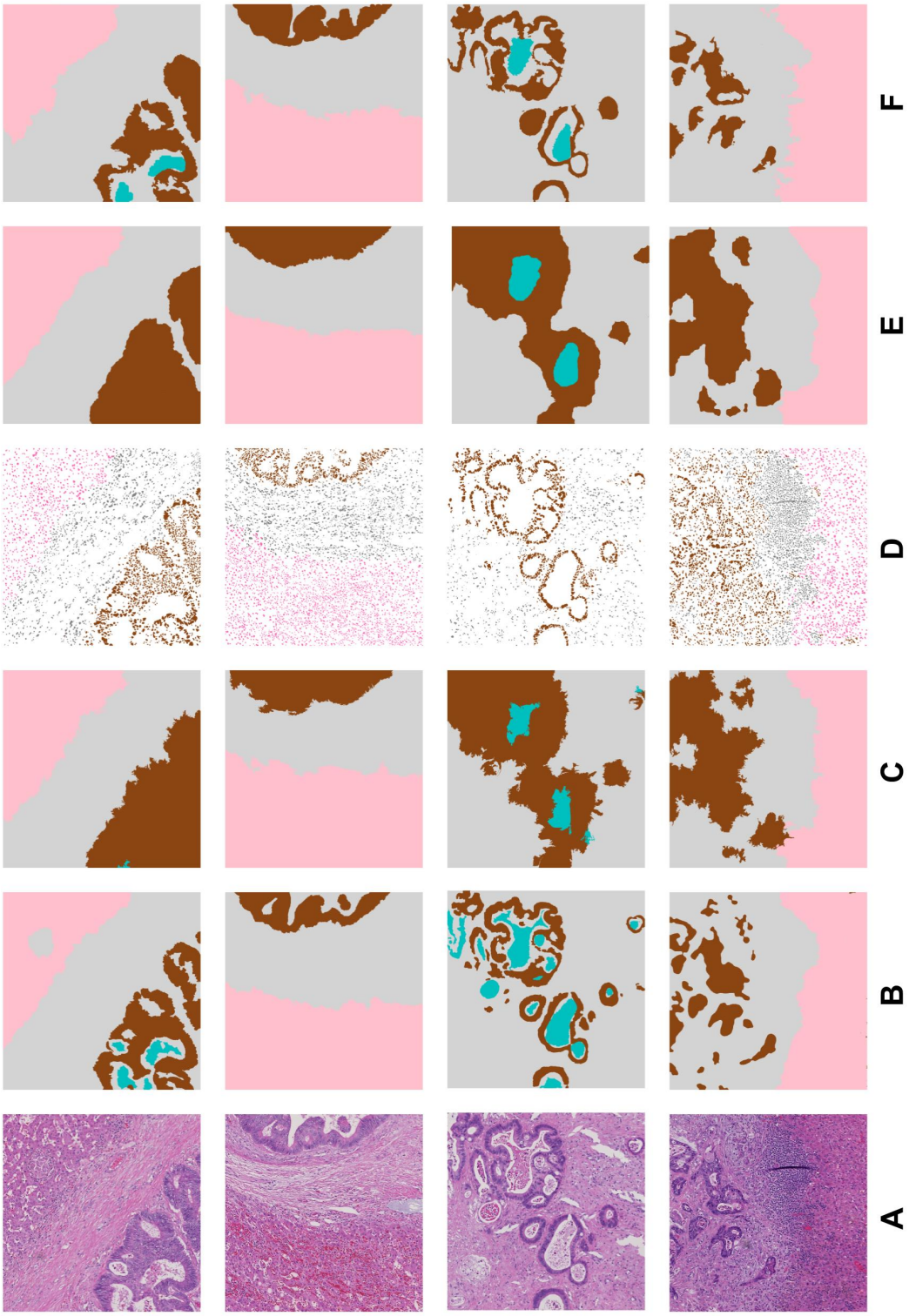


Fig. 5.8 Examples of segmentation results. A : original input image; B: ground truth mask; C: PatchCNN + Superpixel; D: Cell level result; E: U-net with synthesised dataset; F: U-net with transformed dataset.

5.3 Conditional CycleGAN for virtual restaining

In this section, the pixel level segmentation is investigated from the perspective of unpaired image-to-image translation, referred as "virtual restaining". The idea behind is to transform the HE stained image to IHC stained image by correlating the shared structural information with the right colour coding. Then, as long as the image can be correctly transformed, the segmentation result will be easily obtained. A conditional CycleGAN (cCGAN) which could perform multi-class virtual staining with additional patch-wise labelling is proposed. This method extends the original CycleGAN model proposed in [171] for the histology image tasks, by introducing the following two improvements: 1) Patch-wise label information is employed to enforce the model to learn mutual representation with a conditional query. This is achieved by appending a few classifier layers to the original generative network for the sub-domain consistency. 2) Two constraints are introduced to regularize changes to the structural details: the photorealism loss and the structural similarity loss.

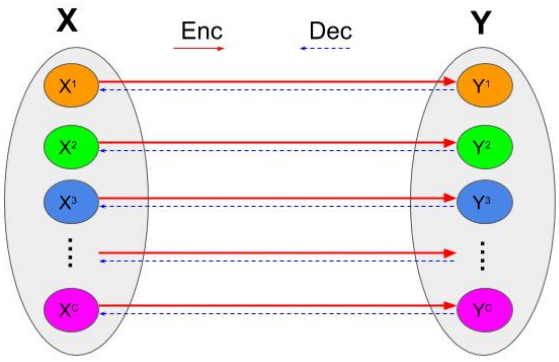


Fig. 5.9 Mapping between two domains and their sub-domains

5.3.1 Proposed method

The overall objective of the proposed network is to learn a multi-class mapping between two domains X and Y. In fact, the transformation takes place between their sub-domains X^t and Y^t, $t \in [1, C]$, where C denotes the number of predefined classes. For two unpaired samples

x and y , their distributions of the training dataset are denoted as $x \sim p(x|c)$ and $y \sim p(y|c)$ where c represents the category condition. As illustrated in Fig. 5.9, there are two conditional mappings in our model, the encoder network Enc maps the input image from domain X to Y while the decoder network Dec maps back from Y to X.

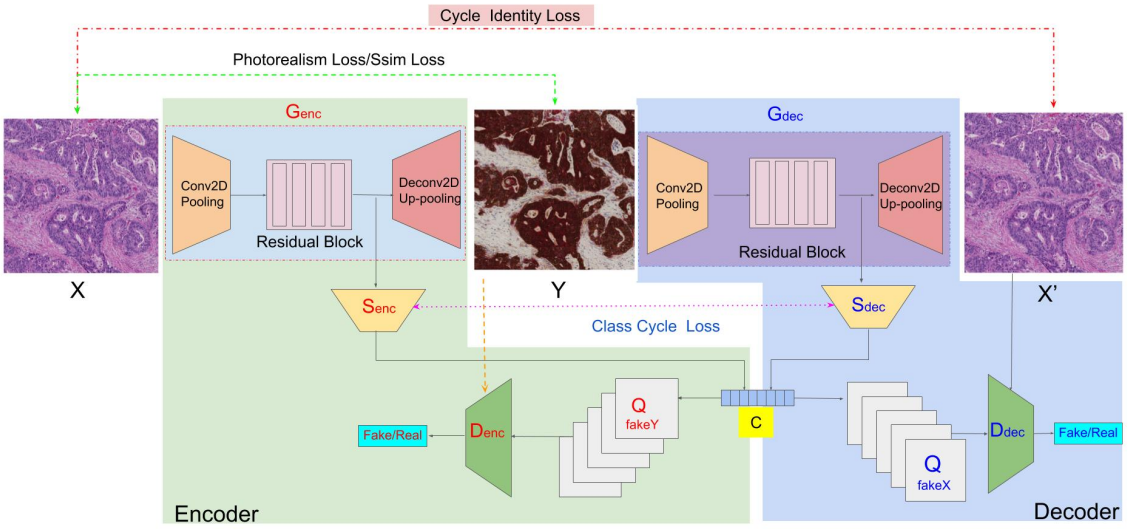


Fig. 5.10 Overall structure of the proposed cCGAN network

As shown in Fig. 5.10, the whole framework consists of two big modules, an encoder network (in green zone) and a decoder network(in blue zone), both of which have the same sub-network structure, a generator G, a discriminator D and a classifier S. In the encoder network, they are denoted as G_{enc}, D_{enc} and S_{enc}. Likewise, in the decoder network, they are G_{dec}, D_{dec} and S_{dec}. The generator is in charge of translating the image between domains, while the discriminator provides feedback on True or False for the generated image. At the same time, the classifiers will provide extra information about the image's sub-domain. To train the proposed model for decent performance on the task, the following losses are facilitated during the training phase.

Conditional adversarial loss

To include the category information for guiding the translation process, conditions are applied on the inputs for the discriminators D_{enc}, D_{dec}. For the mapping from X to Y, the conditional adversarial loss inspired by the loss in [93, 175] within the class c and the generator G_{enc} is defined as :

$$\begin{aligned}\mathscr{L}_{cGAN}(G_{enc},\, D_{enc},\, x,\, y,\, c) = \mathbb{E}_{y\sim p(y|c)}[logD_{enc}(y,\, c)] \\ +\mathbb{E}_{x\sim p(x|c)}[log(1-D_{enc}(G_{enc}(x),\, c))]\end{aligned} \tag{5.11}$$

Likely, for the mapping from Y to X, the same loss is calculated with the discriminator D_{dec} and generator G_{dec}.

Deterministic loss for tissue classification

The tissue sub-type is a kind of crucial information for the histopathology image translation. Without appropriate class inference, it will lead to a failed translation. In our network, the classifiers S infer the category information independently of the generator but share the same base network structure. Practically, by adding extra category information to the network, the objective is to map the distribution of the data category information during the transformation and include the categories information in the output as well. The networks employ the Softmax Entropy Loss ℓ to regularise the category information with regards to their class information. The loss of the classifiers is defined as :

$$\mathscr{L}_{class}(S_{enc},\, S_{dec},\, x,\, y,\, c) = \mathbb{E}_{x\sim p(x|c)}[\ell(S_{enc}(x),\, c)] + \mathbb{E}_{y\sim p(y|c)}[\ell(S_{dec}(x),\, c)] \tag{5.12}$$

Cycle loss and classification cycle loss

The core idea of cycle-GAN is to use the encoder and decoder process as a cycle to make $X \approx X'$ after one cycle $X \rightarrow Y \rightarrow X^{'}$. The cycle structural combines two weak modules and strengthen them using a loop which is vital for the unpaired image translation. By minimising

the difference between X and $X^{'}$, the network will be able to learn the shared features. The original cycle loss defined as [171] :

$$\begin{aligned}\mathscr{L}_{cyc}(G_{enc},\, G_{dec},\, x,\, y) = \mathbb{E}_{x\sim p(x)}[||G_{dec}(G_{enc}(x)) - x||_1] \\ + \mathbb{E}_{y\sim p(y)}[||G_{enc}(G_{dec}(y)) - y||_1]\end{aligned} \tag{5.13}$$

During the cycle process, the outputs of the classifiers S should be identical as well. Thus the classification cycle loss is defined with L1 loss:

$$\mathscr{L}_{clcyc}(S_{enc}, S_{dec}, x, y) = \mathbb{E}_{(x,y)\sim p(x,y|c)}[||S_{enc}(x) - S_{dec}(y)||_1] \tag{5.14}$$

The cycle consistency of the classification information is of signification importance to final results as well.

Identity loss, photorealism loss and structural similarity Loss

To rectify the output in the middle of the process is in the right track, three more loss functions are employed to ensure the structure information unchanged and to enhance the original cycle-GAN network. The identity loss and photorealism loss, introduced in Sec. 5.2, for the proposed network are:

$$\mathscr{L}_{idt}(G_{enc},\, G_{dec},\, x,\, y) = \mathbb{E}_{x\sim p(x)}[||(G_{enc}(y) - x||_1] + \mathbb{E}_{y\sim p(y)}[||G_{dec}(y) - y||_1] \tag{5.15}$$

$$\mathscr{L}_{pho}(G_{enc},\, G_{dec},,\, x,\, y) = Pho(G_{enc}(x),\, x) + Pho(G_{dec}(y),\, y) \tag{5.16}$$

SSIM has been used for assessing the image quality in many related studies [176]. In this task, for each pixel in the image, the SSIM is defined as:

$$Ssim(x,y) = \frac{2\mu_x\mu_y + Q_1}{\mu_x^2 + \mu_y^2 + Q_1} + \frac{2\sigma_{xy} + Q_2}{\sigma_x^2 + \sigma_y^2 + Q_2} \tag{5.17}$$

where $\mu_x,\ \mu_y$ are the mean of a fixed window centred as the pixel, $\sigma_x,\ \sigma_y$ are the standard derivations. $Q_1,\ Q_2$ are the regularisation term for division stabilisation. Hence, the loss function of SSIM of the whole network can be formulated as:

$$\mathscr{L}_{sim}(G_{enc},\ G_{dec},\ x,\ y) = (1 - Ssim(G_{enc}(x),\ x)) + (1 - Ssim(G_{dec}(y),\ y)) \tag{5.18}$$

Proposed Approach

Our full objective can be achieved by a weighted linear combination of different losses:

$$\begin{aligned}\mathscr{L}(G,\ D,\ S,\ x,\ y,\ c) = \mathscr{L}_{cGAN}(G_{enc},\ D_{enc},\ x,\ y,\ c) + \mathscr{L}_{cGAN}(G_{dec},\ D_{dec},\ x,\ y,\ c)\\ +\lambda\mathscr{L}_{cyc}(G_{enc},\ G_{dec},\ x,\ y) + \delta\mathscr{L}_{id}(G_{enc},\ G_{dec},\ x,\ y)\\ +\gamma\mathscr{L}_{class}(S_{enc},\ S_{dec},\ x,\ y,\ c) + \gamma\mathscr{L}_{clcyc}(S_{enc},\ S_{dec},\ x,\ y)\\ +\alpha\mathscr{L}_{ssim}(G_{enc},\ G_{dec},\ x,\ y) + \beta\mathscr{L}_{pho}(G_{enc},\ G_{dec},\ x,\ y)\end{aligned} \tag{5.19}$$

The parameters $\lambda,\ \gamma,\ \delta,\ \alpha,\ \beta$ in the loss function regulate the importance of different losses to the overall objective. By solving the following equation,

$$G,\ D,\ S = \arg\min_{G,\ S}\max_{D}\mathscr{L}(G,\ D,\ S,\ x,\ y,\ c) \tag{5.20}$$

optimal models can be found for the generators $G_{enc},\ G_{dec}$, the decoders $D_{enc},\ D_{dec}$ and the classifiers $S_{enc},\ S_{dec}$.

5.3.2 Dataset and implementation

Dataset preparation

The WSIs used for this task is from the same collection as explained in Sec. 5.2. However, as stated before, the virtual re-staining task is more challenging than the virtual registration task. The different datasets are used in this section. As demonstrated in Fig. 5.11, they

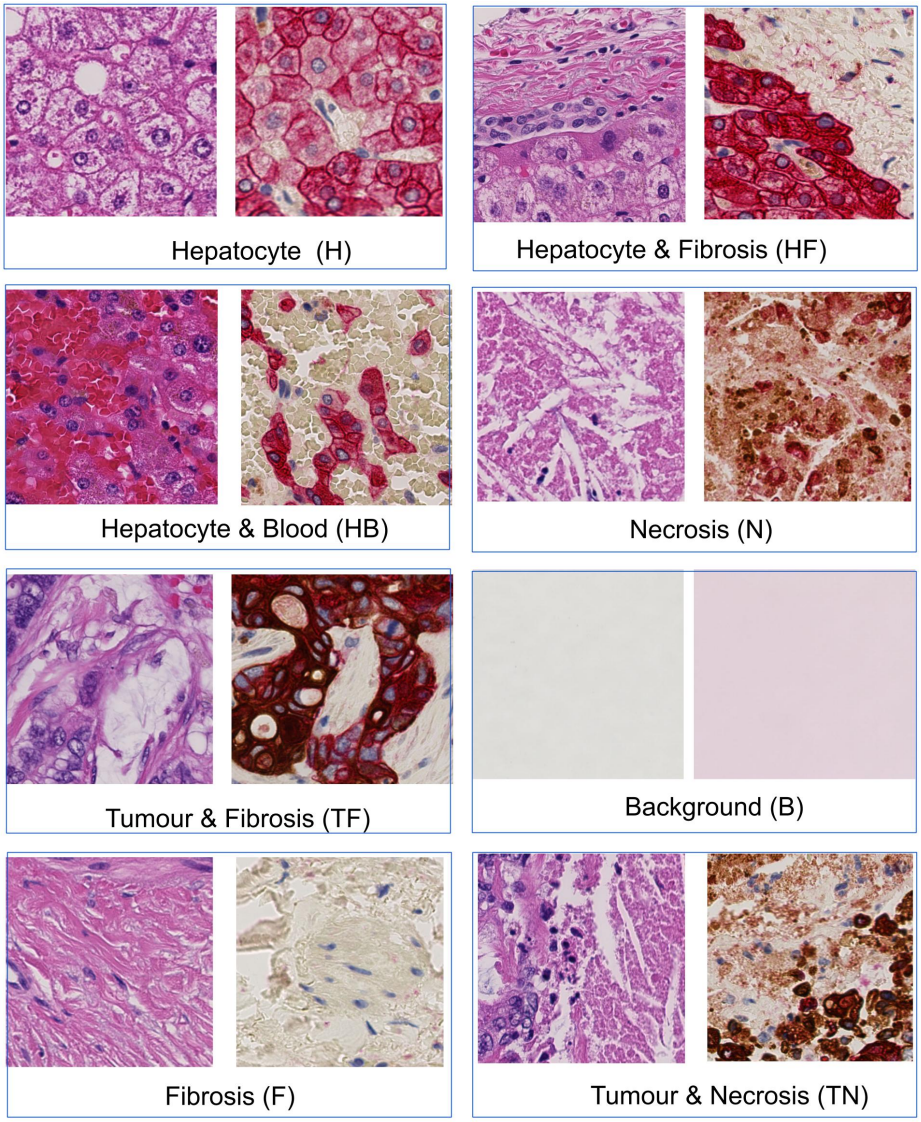


Fig. 5.11 Example of the unpaired training dataset

are *Hepatocyte (H), Fibrosis (F), Necrosis (N), Tumour & Fibrosis (TF), Hepatocyte & Fibrosis (HF), Hepatocyte & Blood (HB), Tumour & Necrosis (TN) and Background (BG).* The training dataset is cropped from aligned WSIs with a patch size of 256×256 pixels on $20\times$ magnification. The number of patches that are used for training is listed in Table 5.2.

Network Architecture This section describes the implementation details including the network structure and the parameters set-up. The generator networks have the same architecture as proposed in [177]. Two convolutional layers at the beginning for down-sampling and

two deconvolutional layers for up-sampling. In the middle, there are 9 Resnet blocks [138]. Both the discriminators and classifiers are composed in a fully convolutional fashion. For the discriminators, there are 5 layers inside, while for the classifiers, the number of convolutional layers is 8.

Training Details During the training, the history of generated images is used to reduce model oscillation. However, to fit with the conditional generative network, the image query process is applied on condition as well. For the parameters $\lambda, \gamma, \delta, \alpha, \beta$ in the loss functions Eq.5.19, λ is set to a fixed value of 10, γ is set to 0.5 and the other three (δ, α, β) are set to different values to compare the performances. The rest of the parameters are identical with the original CycleGAN.

5.3.3 Results and discussion

Evaluating the results of the generated images for tasks like style transfer and image re-colourisation is a well-known challenge for a long time. For the virtual staining tasks, two aspects should be considered. The semantic meaning of each re-colourised regions and the structure/content information. The semantic meaning refers to the colour coding of the virtual stained and IHC stained should be the same and it denotes how well the model can translate the original images. It would be very definitely wrong if a hepatocyte area is virtually stained with brown which is the colour of the tumour area. The structure information relates to the location and texture of the pixel in the tissue patches. If the location of the nucleus or the existence of a certain nucleus is changed, it means the result is not satisfying.

Subjective evaluation

One of the wide adopted methods is to use Amazon Mechanical Turk (AMT) perceptual study that asks human observers to differentiate the fake from the true ones. However, in our study, well-trained histopathology experts are hired to distinguish them. Meanwhile, the

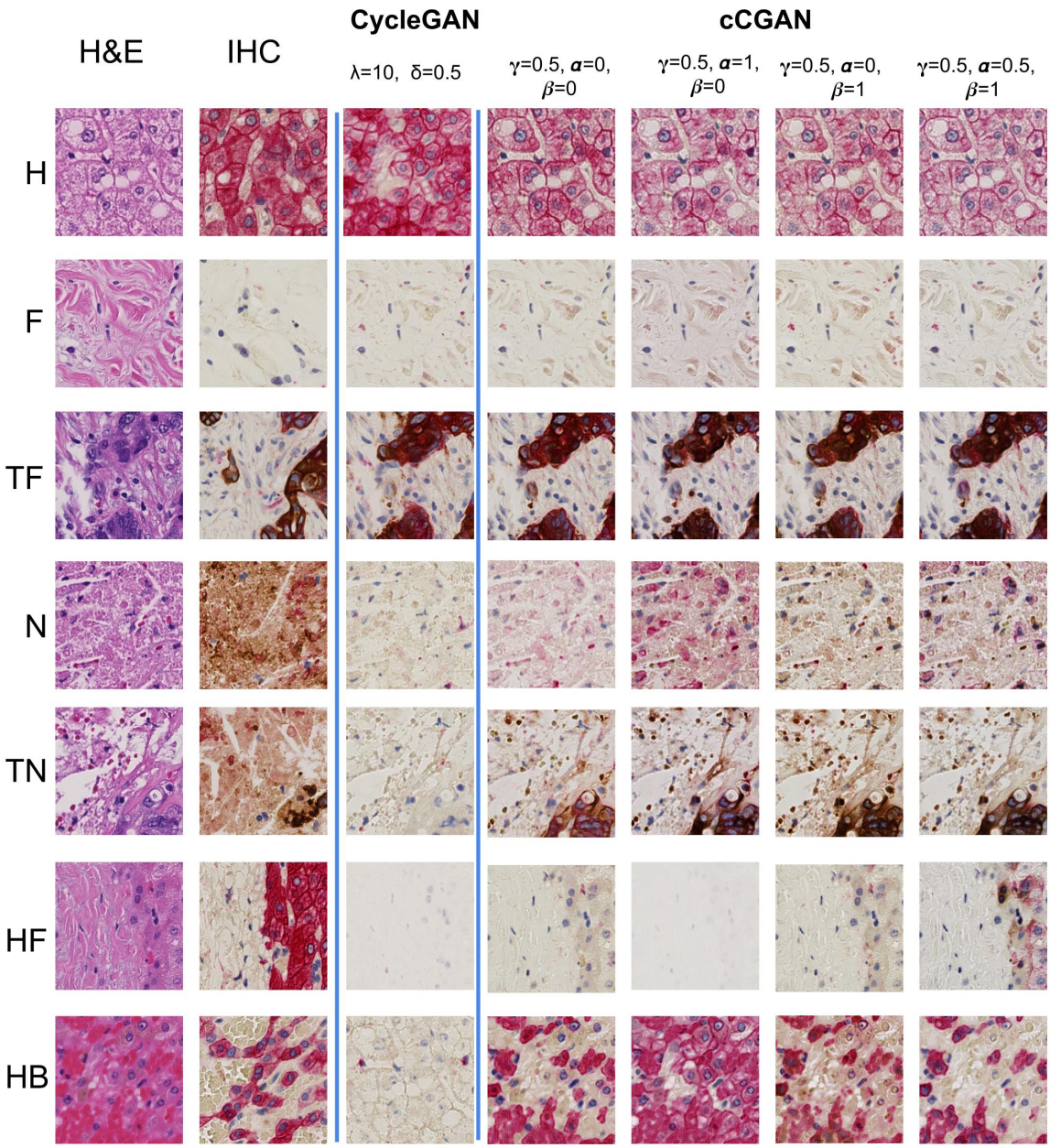


Fig. 5.12 Proposed cCGAN model results with different parameters settings

computer vision experts are asked to assess on the image by a visual quality error score. For the computer vision researchers, all the patches (real and fake) from different classes are mixed together and the observers need to make decisions based on the image quality. While, for professional assessment, both of the original HE stained and virtual stained patches will be provided. The experts will give a staining score that indicates if the patches have been properly virtual stained. Two pathology experts and ten computer vision researchers are invited to perform the evaluation. 240 patches are generated from the evaluation dataset (30 patches per class). The results of experts are listed in Table 5.2. The results of different settings are demonstrated in Fig. 5.12. In this figure, for each HE stained patch, the patch from the same location on the roughly aligned IHC images are used as references. Comparing to the results by CycleGAN, the results of the proposed method greatly improve the results especially in the class with mixed components and limited training examples such as *TN, HF, HB*. From the experiments, it can be observed the $\gamma = 0.5, \alpha = 0.5$ and $\beta = 1$ output the best results. Hence, the results from this model are employed for further evaluations. Further results are listed in Table 5.2.

Table 5.2 Quantitative evaluation results of the models for virtual restaining

	H	TF	N	F	HF	TN	HB	BG	All
No. Annotated patches (HE)	1828	950	538	798	212	48	394	2016	3392
No. Annotated patches(IHC)	1440	738	438	1184	210	52	88	1252	2711
No. Test patches (HE)	30	30	30	30	30	30	30	30	240
Vision experts on CycleGAN*	3%	26%	1%	4%	3%	0%	18%	0%	7%
Vision experts on cCGAN*	0%	10%	4%	0%	0%	0%	0%	0%	2%
Pathologists on CycleGAN*	34%	71%	22%	11%	83%	95%	97%	0%	52%
Pathologists on cCGAN*	23%	39%	23%	4%	76%	86%	99%	0%	44%

* The number indicates the visual quality error rate and staining error rate. The larger the number indicate a worse performance.

The results from the non-experts view demonstrated that the proposed method is better to preserve the image contents than the original CycleGAN. The introduction of the structural losses to the model greatly suppresses the "imaginary" ability, especially when encountering new features. The class *TF* has the highest fake rate, which has a strong relationship with its complex features. The poor performance in the classes *HB, HF, and TN* is due to the insufficient and unbalanced training data in different classes. By inspecting the failed transformation cases, it can be observed that mapping to the wrong domain is the main cause for the failure. Some examples are illustrated in Fig. 5.13. The patch from H is mapped to F, while for the HB, N, and TN, the staining style is N, F and H respectively.

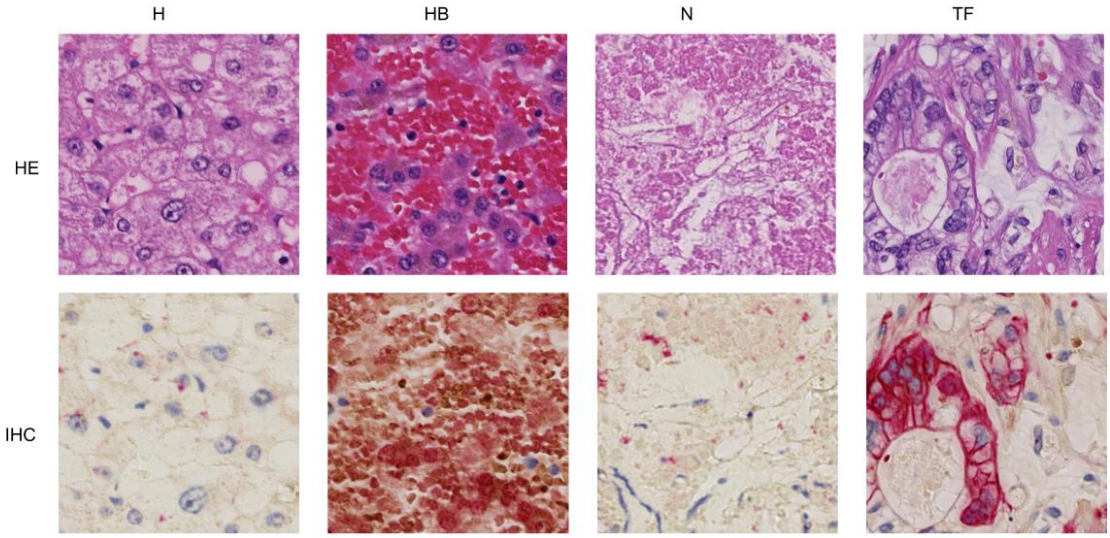


Fig. 5.13 Unsuccessful mapping examples of the virtual restaining model

Visual assessment

Visual assessment is one of the most straightforward way to assess the results from the virtual staining. Instead of demonstrating results from the testing dataset as in Fig. 5.12, the virtual staining results on large patches ($1024px \times 1024px$ on $20\times$ magnification level) and on the WSIs are demonstrated in Fig. 5.14 and Fig. 5.15.

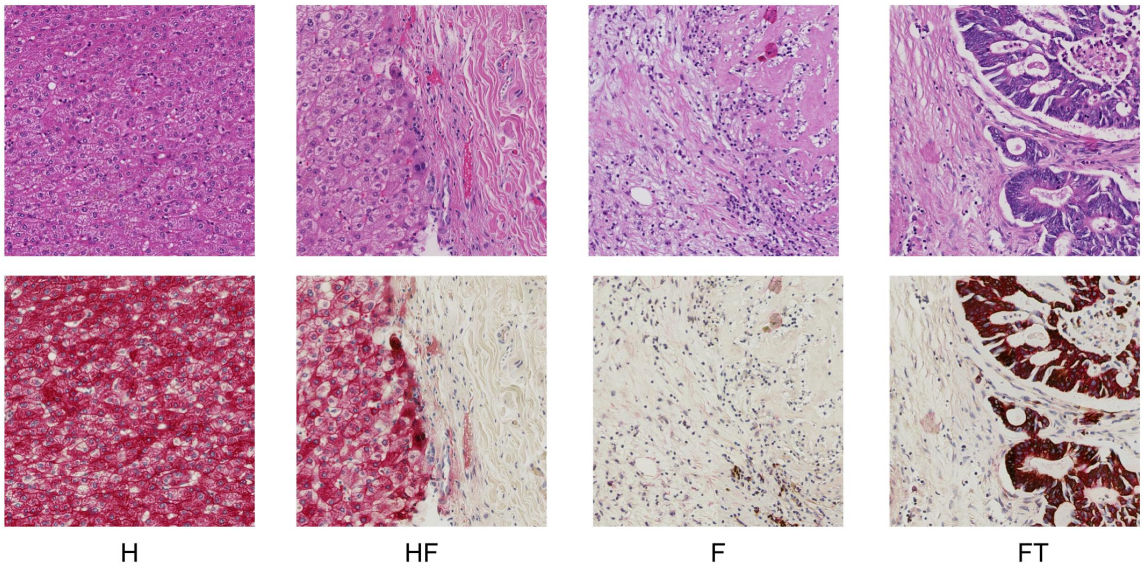


Fig. 5.14 Virtual restaining results on the large patches

In Fig. 5.14, patches with different morphologies (H, HF, F, FT) are demonstrated, the first row is the original HE stained patches, the second row is the virtual restained images. In Fig. 5.15, there are 4 thumbnails of the WSIs. (A) and (D) are the physical stained images, (B) and (C) are the virtually stained results. Hence, from (A) to (B) is the virtual restaining process, from (D) to (C) is the virtual registration process. Although, (A) and (D) look similar to each other, they are not exactly the same slide, (D) is actually a neighbouring slide of (A). Both of the results are obtained with the cCGAN model ($\rho = 0.5, \alpha = 0.5$ *and* $\beta = 1$) using sliding window method, the window size set as $256px \times 256px$. However, the results on the patch level have been smoothed, which means there are overlaps between different windows and the final output is the average results on multiple windows. While the results on the WSIs have no post-processing operations, the window size is equal to the patch size, that is, simply stitching patches together. As it can be observed, with smoothing, the results are very consistent on the large patch; there is no obvious edge between the small tiles. Conversely, the images from Fig. 5.15 (B, C) have mosaic on them. The edges between different tiles make the image have a noisy overall appearance. Regardless of the artefacts, the results on the WSIs reveals that the model has good stability and excellent generalisability.

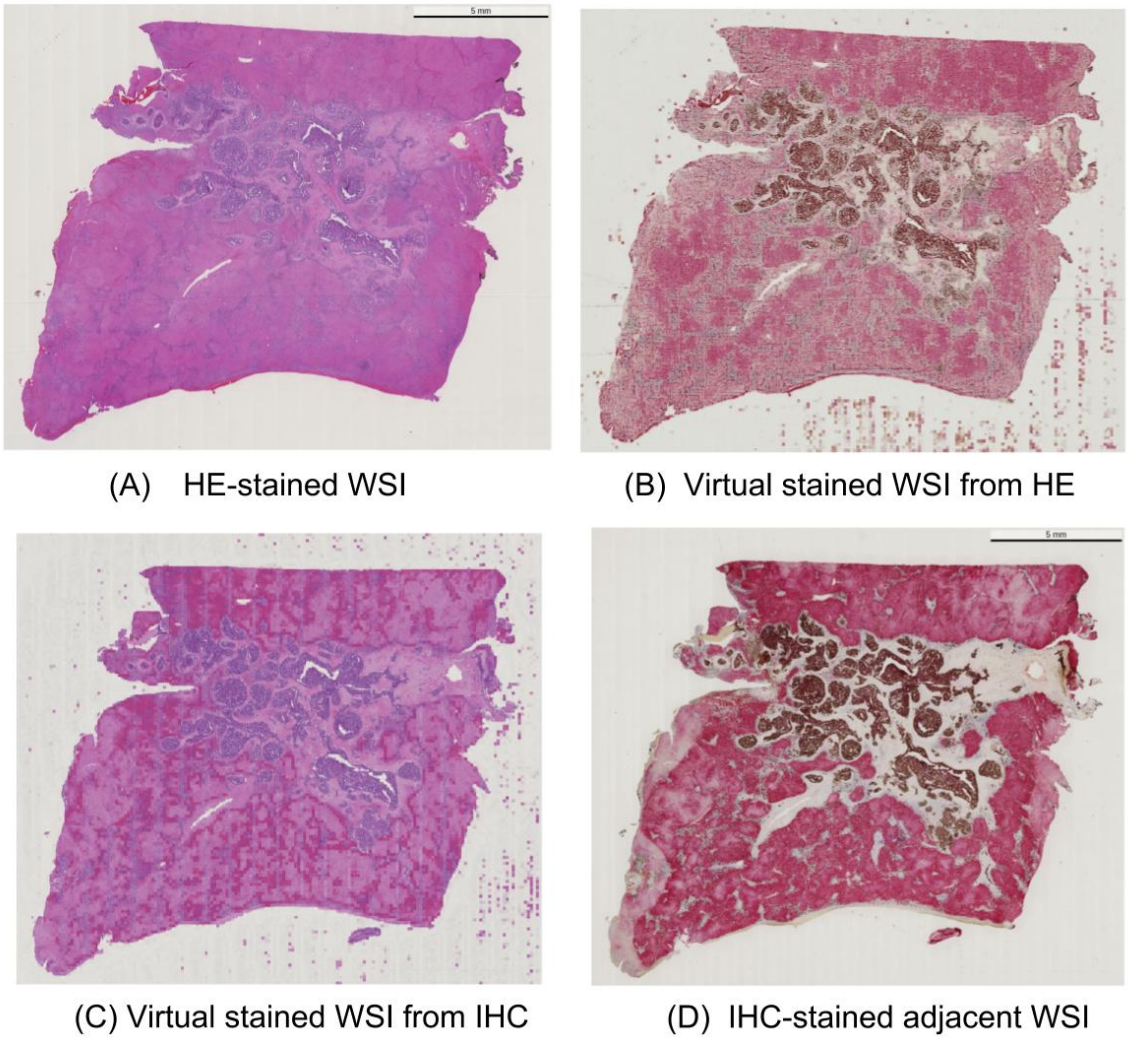


Fig. 5.15 Virtual restaining and virtual registration results on the whole slide image

Objective Evaluation

The pixel level segmentation can be considered as a special kind of virtual staining, vice versa, the virtual staining can be evaluated with segmentation metric as well. To evaluate the results by virtual restaining, the 50 ROIs for assessing different segmentation approaches are virtually stained. Then the virtual stained images will be processed with the same pipeline as mentioned in the last section, where the IHC stained images are used to generate HE stained images and masks. The final quantitative results are listed in Table 5.3. However, the

evaluation method for this objective evaluation is different from the previous ones since the necrosis is not considered, which can be observed from Fig. 5.18.

Table 5.3 Segmentation results (PA & MIoU) based on virtual restained image

	CycleGAN (a)	cCGAN (b)
PA	83.71%	88.49%
MIoU	61.82%	65.76%

(a): Virtual stained results with cycle GAN model. (b): Virtual stained result with cCGAN model. Same post-processing method applided on (c) and (d)

The quantitative results in the table indicate that the objective of virtual staining and pixel level segmentation are identical. Both of them tend to analysis the histopathology image based on pixel level information. Besides, it also can be observed that the proposed cCGAN model performs better than the original cycle GAN model in the segmentation task, which can be considered as a subjective evaluation of virtual restaining task. However, the results based on the virtual restaining is not as good as the results from the model using supervised training methods. The main possible reason is that the ground truth from the experts is mainly from the semantic perspective, which is an approximation to the real border. While, the virtual staining result will try to locate the real edge and the details, even though it can be noisy.

Discussion The results indicate that the proposed virtual staining method is capable of virtually staining the HE stained images into IHC stained style image based on the underlying mutual features in the two kinds of images. In other words, IHC stained WSI are distinguishable enough for guiding the cross-domain translation.

With the help of virtual staining, the laborious and expensive experts' manual annotation can be minimised. In fact, the presented method for computerised virtual staining entails great potential for a wide range of clinical diagnostic and AI related applications. A few of these are listed below:

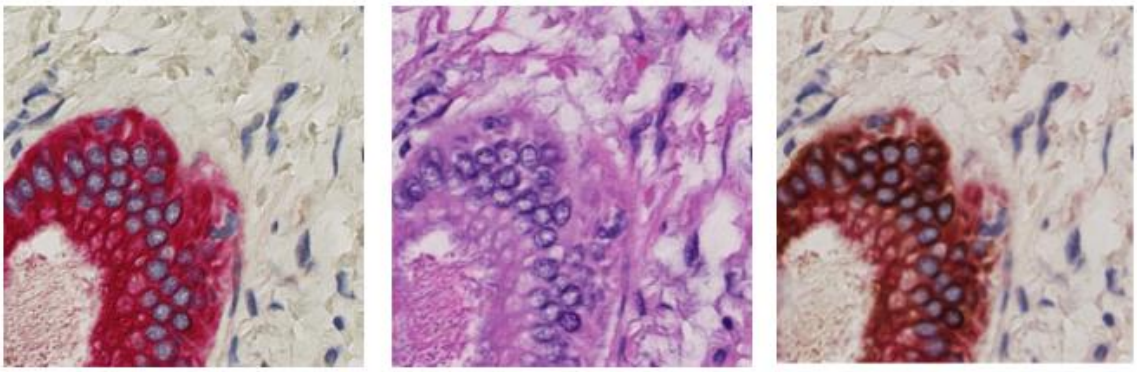


Fig. 5.16 Example results for image registration and multipixel. The three images are from the same region but with different stainings, IHC (CD7/CD20), HE, IHC (CD18/CD19), from left to right.

1. Fast and low-cost generation of IHC staining for improved tumour diagnosis. Currently, pathologists devote a great amount of time and efforts in the error-prone eyeball work of searching for tumour cells within the low-contrast HE slides. Virtual IHC of tumour cells, by enabling high-contrast visualisation, may contribute to greatly reducing visual screening time and improving the accuracy of cancer detection.

2. Virtual multiplex IHC staining for spatial mapping and objective quantification of intertwined cell types. As a generalisation, GAN based model can be trained to classify and virtually stain multiple cell types of interest, provided virtual multiplex immunostainings. Fig. 5.16 demonstrates a potential use case of the virtual multiplex.

3. WSI registration for colocation analysis in multiple IHC stainings. Registration of multiple serial sections is another challenge of increasing interest in the diagnosis and biomedical research. In clinical practice, the pathologists employ multiple IHC stainings (biomarkers) to detect relevant features in cells and tissues within the WSIs. As it is often not feasible to align the samples during lab processing, image registration is the only option to accomplish this task, but very difficult to achieve. By exploiting virtual staining's high accurate matching of image pairs, multiple aligned IHC stainings

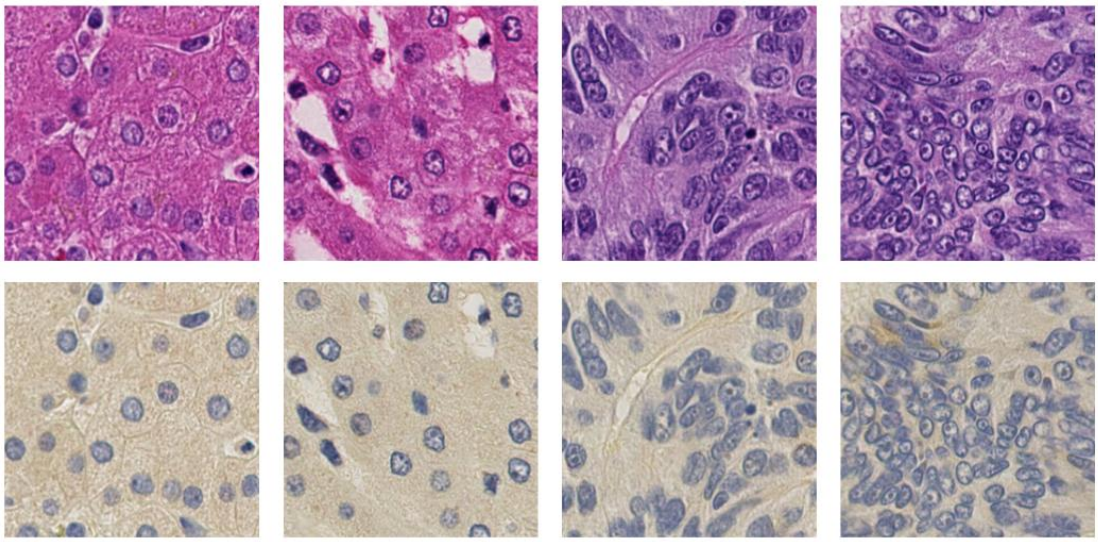


Fig. 5.17 Colour deconvolution examples using virtual staining framework

can be combined into a new staining style achieving increasing levels of multiplex immunostaining far beyond current laboratory techniques.

4. Color deconvolution and nuclei detection. Virtual staining can also be used to decouple the information of the HE stained image to obtain only the H channel of the image. By using HE, H-only, and E-only stained WSIs as training data, the model will learn to "unmix" the two colour components in the HE images. The H/E-only images can then be easily normalised. This may facilitate nuclei detection tasks, especially in cell dense and cancerous areas. Some preliminary results are demonstrated in Fig. 5.17.

One of the limitations of the current study is the low number of samples for training and the significant imbalance on the amount of annotated patches among different classes. In future work, attempts will be paid to minimise or discarding the labelling information that is used as input in order to pursue unsupervised image translation.

5.4 Summary

In this chapter, the histopathology image analysis is processed from another perspective - virtual staining. With the presence of the IHC stained image, the analysis of HE stained image can be converted a virtual staining problem. The virtual staining framework, which can be treated as a bridge connecting different tasks. To this end, it could provide a unique and universal solution for different histopathology image analysis problems.

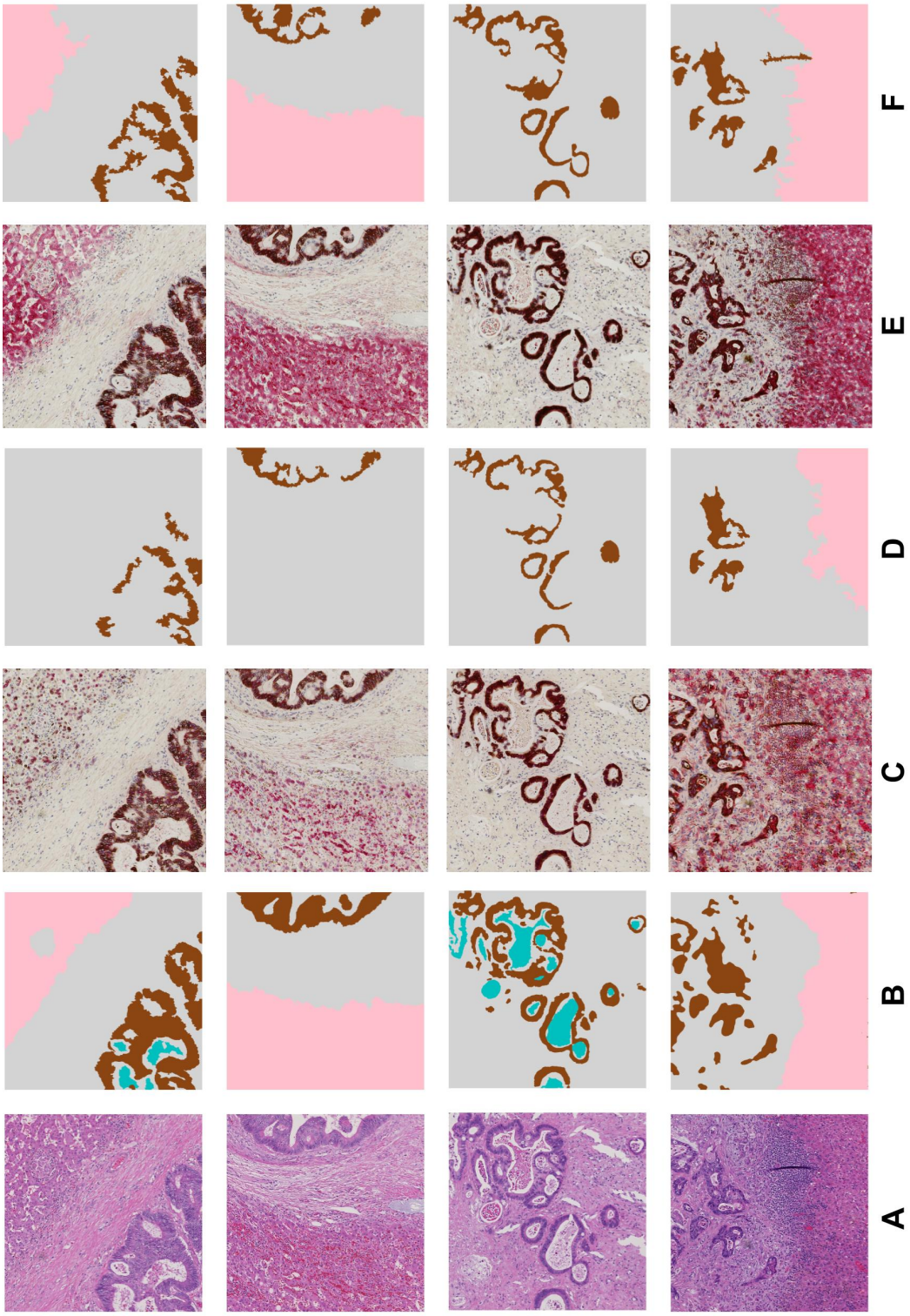


Fig. 5.18 Examples of virtual staining results on the testing dataset. A : original input image; B: ground truth mask; C: Virtual stained result with CycleGAN; D: Mask obtained with CycleGAN; E: Vitual-stained result wiht cCGAN; F: Mask obtained with cCGAN.

Chapter 6

Conclusion

This thesis investigates the problem of quantitative analysis of histopathology images. Several algorithms or frameworks are proposed to enhance the results in different tasks and address the challenges of insufficient data annotations. In this chapter, the findings and limitations identified in the previous chapters are summarised.

6.1 Findings and Limitations

The findings and limitations of this thesis can be summarised from three aspects, which correspond to the three main chapters.

In Chapter 3, patch-based tissue type classification and segmentation are explored. Different networks are evaluated with different patch size and magnification levels. Extensive experiments reveal that a larger patch size will lead to better performance due to more contextual information being included. An approximate segmentation result through the sliding window method can be obtained, but the limitation of the classification based results is that the results are of low resolution. Furthermore, a PatchCNN model is proposed to enhance the resolution of the output through data synthesis. Then, the segmentation based method is implemented to further enhance the segmentation results. The limitation of both

the classification based and segmentation based methods is that the algorithms are not very sensitive to small tumour regions.

Chapter 4 sought to facilitate cell level information for tissue segmentation since the numerous nuclei inside the WSI can be clearly observed. The findings described in this chapter can be considered from two aspects. From the cell level, this research confirms that it is possible to classify the nuclei with weakly supervised methods. In terms of detection and segmentation, the experiments here demonstrate that it is possible to transfer the model between different kinds of dataset. For the fusion model, the main findings are that the morphology and location information can help to refine the border between different tissues. To this end, the insufficient data annotation problem can be solved by using the abundant information at the cell level. The limitation of this method is the long processing time for large patches.

Transformation is the main guideline of Chapter 5. The conception of virtual restaining, virtual registration and virtual magnification are proposed. The pixel level image segmentation problem is solved through the transformation between HE stained and IHC stained images. The high contrast feature of the IHC images can help to generate pixel-level masks which are essential for supervised learning. Through the virtual registration, the pairs of generated HE stained images and their corresponding masks can acquired, the model is further trained to conduct the semantic segmentation task. The virtual restaining, whereas, can be considered as an auxiliary visualisation method. However, the limitation for the promising application on the histopathology images is the feasibility to make a qualified evaluation.

6.2 Future Work

In this research, most of the research has been conducted from a computer vision researcher's point of view. More work should be conducted with close collaboration from the clinical side. Further ideas for future research are highlighted as follows:

Link with clinical data. The main purpose of the histopathology image analysis is to assist the pathologists, reducing their workload or improving the work quality. In the context histopathological image analysis, the images are an important reflection of disease progression. Together with the survival data, the quantitative analysis of the tissue components in the WSI can be used to predict the progress of the disease. Furthermore, by analysing the survival data, it is possible to identify the related features that are mostly likely to determine the progression of disease, such as the patterns of the invasive front.

Super virtual staining. The idea of super-resolution is to solve all the histopathological image analysis related tasks in one model based on the virtual re-staining framework. One of the most challenging problems of most medical related image analysis is the difficulty in obtaining sufficient annotations. However, if there are opportunities to encourage the model to transform different types of images based on structural features and high-level semantic features, the problems that need supervised training will be transferred to weakly supervised methods.

3D histology image analysis. The nature of the histopathology image of the specimens is based on the cross sections of 3D objects. This is the same for the cell or nuclei as well. Hence, there should be more information or multi-dimensional information to acquire from the 3D objects, which means that 2D section images will contain some biases for analysis. If it were possible to combine multiple slides to reconstruct the 3D object, the growing pattern of the tissue could be analysed in 3D dimensions. Furthermore, the model can provide a more dedicated 3D model of the objects, illustrating their inner structure.

Integrate with genomics data. Genome analysis is a common technique in biomedical related research. In clinical research, genomics data are closely related to the clinical index, which can provide another source to support disease diagnosis from another perspective. Both can be considered a part of the "micro-environment" since the visual appearance or the structure of the tissue in histopathology images is a result of gene expression. In this case,

the mechanisms of how the disease develops and the signal that can be used for diagnosis are heated research topics. With the combination of these two kinds of data, especially if they can be linked, this will lead to a great breakthrough in the biomedical research field.

References

[1] Michael T McCann, John A Ozolek, Carlos A Castro, Bahram Parvin, and Jelena Kovacevic. Automated histology analysis: Opportunities for signal processing. *IEEE Signal Processing Magazine*, 32(1):78–87, 2015.

[2] Thomas J Fuchs and Joachim M Buhmann. Computational pathology: Challenges and promises for tissue analysis. *Computerized Medical Imaging and Graphics*, 35(7):515–530, 2011.

[3] Shaimaa Al-Janabi, André Huisman, and Paul J Van Diest. Digital pathology: current status and future perspectives. *Histopathology*, 61(1):1–9, 2012.

[4] John Arevalo, Angel Cruz-Roa, et al. Histopathology image representation for automatic analysis: A state-of-the-art review. *Revista Med*, 22(2):79–91, 2014.

[5] Lindsey A Torre, Freddie Bray, Rebecca L Siegel, Jacques Ferlay, Joannie Lortet-Tieulent, and Ahmedin Jemal. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*, 65(2):87–108, 2015.

[6] Rebecca L Siegel, Kimberly D Miller, Stacey A Fedewa, Dennis J Ahnen, Reinier GS Meester, Afsaneh Barzi, and Ahmedin Jemal. Colorectal cancer statistics, 2017. *CA: a cancer journal for clinicians*, 67(3):177–193, 2017.

[7] Ching-Wei D Tzeng and Thomas A Aloia. Colorectal liver metastases. *Journal of Gastrointestinal Surgery*, 17(1):195–202, 2013.

[8] Antoine Brouquet, Eddie K Abdalla, Scott Kopetz, Christopher R Garrett, Michael J Overman, Cathy Eng, Andreas Andreou, Evelyne M Loyer, David C Madoff, Steven A Curley, et al. High survival rate after two-stage resection of advanced colorectal liver metastases: response-based selection and complete resection define outcome. *Journal of Clinical Oncology*, 29(8):1083–1090, 2011.

[9] Christophe Penna and Bernard Nordlinger. Surgery of liver metastases from colorectal cancer: new promises. *British medical bulletin*, 64(1):127–140, 2002.

[10] Inti Zlobec and Alessandro Lugli. Invasive front of colorectal cancer: dynamic interface of pro-/anti-tumor factors. *World journal of gastroenterology: WJG*, 15(47):5898, 2009.

[11] Rikke Løvendahl Eefsen, Gert G Van den Eynden, Gunilla Høyer-Hansen, Pnina Brodt, Ole Didrik Laerum, Peter B Vermeulen, Ib Jarle Christensen, André Wettergren, Birgitte Federspiel, Gro L Willemoe, et al. Histopathological growth pattern,

proteolysis and angiogenesis in chemonaive patients resected for multiple colorectal liver metastases. *Journal of oncology*, 2012, 2012.

[12] PH Bartels, JE Weber, and L Duckstein. Machine learning in quantitative histopathology. *Analytical and quantitative cytology and histology*, 10(4):299–306, 1988.

[13] Juan C Caicedo, Angel Cruz, and Fabio A Gonzalez. Histopathology image classification using bag of features and kernel functions. In *Artificial intelligence in medicine*, pages 126–135. Springer, 2009.

[14] Angel Cruz-Roa, Juan C Caicedo, and Fabio A González. Visual pattern mining in histology image collections using bag of features. *Artificial intelligence in medicine*, 52(2):91–106, 2011.

[15] Mutlu Mete and Umit Topaloglu. Statistical comparison of color model-classifier pairs in hematoxylin and eosin stained histological images. In *Computational Intelligence in Bioinformatics and Computational Biology, 2009. CIBCB'09. IEEE Symposium on*, pages 284–291. IEEE, 2009.

[16] Doaa Mahmoud-Ghoneim. Optimizing automated characterization of liver fibrosis histological images by investigating color spaces at different resolutions. *Theor Biol Med Model*, 8:25, 2011.

[17] Matthew D DiFranco, Gillian O'Hurley, Elaine W Kay, R William G Watson, and Padraig Cunningham. Ensemble based system for whole-slide prostate cancer probability mapping using color texture features. *Computerized medical imaging and graphics*, 35(7):629–645, 2011.

[18] Rachel Sparks and Anant Madabhushi. Content-based image retrieval utilizing explicit shape descriptors: applications to breast mri and prostate histopathology. In *SPIE Medical Imaging*, pages 79621I–79621I. International Society for Optics and Photonics, 2011.

[19] Kien Nguyen, Bikash Sabata, and Anil K Jain. Prostate cancer grading: Gland segmentation and structural features. *Pattern Recognition Letters*, 33(7):951–961, 2012.

[20] James P Monaco, John E Tomaszewski, Michael D Feldman, Ian Hagemann, Mehdi Moradi, Parvin Mousavi, Alexander Boag, Chris Davidson, Purang Abolmaesumi, and Anant Madabhushi. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models. *Medical image analysis*, 14(4):617–629, 2010.

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

[22] Angel Cruz-Roa, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, and Anant Madabhushi.

Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In *SPIE medical imaging*, pages 904103–904103. International Society for Optics and Photonics, 2014.

[23] Geert Litjens, Clara I Sánchez, Nadya Timofeeva, Meyke Hermsen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-Van De Kaa, Peter Bult, Bram Van Ginneken, and Jeroen Van Der Laak. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports*, 6:26286, 2016.

[24] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*, 2016.

[25] Yun Liu, Krishna Gadepalli, Mohammad Norouzi, George E Dahl, Timo Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Philip Q Nelson, Greg S Corrado, et al. Detecting cancer metastases on gigapixel pathology images. *arXiv preprint arXiv:1703.02442*, 2017.

[26] Yi Li and Wei Ping. Cancer metastasis detection with neural conditional random field. *arXiv preprint arXiv:1806.07064*, 2018.

[27] 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.

[28] Pierre Courtiol, Eric W Tramel, Marc Sanselme, and Gilles Wainrib. Classification and disease localization in histopathology using only global labels: A weakly-supervised approach. *arXiv preprint arXiv:1802.02212*, 2018.

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

[30] Matthias Kohl, Christoph Walz, Florian Ludwig, Stefan Braunewell, and Maximilian Baust. Assessment of breast cancer histology using densely connected convolutional networks. In *International Conference Image Analysis and Recognition*, pages 903–913. Springer, 2018.

[31] Teresa Araújo, Guilherme Aresta, Eduardo Castro, José Rouco, Paulo Aguiar, Catarina Eloy, António Polónia, and Aurélio Campilho. Classification of breast cancer histology images using convolutional neural networks. *PloS one*, 12(6):e0177544, 2017.

[32] Kamyar Nazeri, Azad Aminpour, and Mehran Ebrahimi. Two-stage convolutional neural network for breast cancer histology image classification. In *International Conference Image Analysis and Recognition*, pages 717–726. Springer, 2018.

[33] Yeeleng S Vang, Zhen Chen, and Xiaohui Xie. Deep learning framework for multi-class breast cancer histology image classification. In *International Conference Image Analysis and Recognition*, pages 914–922. Springer, 2018.

[34] Sulaiman Vesal, Nishant Ravikumar, AmirAbbas Davari, Stephan Ellmann, and Andreas Maier. Classification of breast cancer histology images using transfer learning. In *International Conference Image Analysis and Recognition*, pages 812–819. Springer, 2018.

[35] Xinpeng Xie, Yuexiang Li, and Linlin Shen. Active learning for breast cancer identification. *arXiv preprint arXiv:1804.06670*, 2018.

[36] Ruqayya Awan, Navid Alemi Koohbanani, Muhammad Shaban, Anna Lisowska, and Nasir Rajpoot. Context-aware learning using transferable features for classification of breast cancer histology images. In *International Conference Image Analysis and Recognition*, pages 788–795. Springer, 2018.

[37] Alexander Rakhlin, Alexey Shvets, Vladimir Iglovikov, and Alexandr A Kalinin. Deep convolutional neural networks for breast cancer histology image analysis. In *International Conference Image Analysis and Recognition*, pages 737–744. Springer, 2018.

[38] Aditya Golatkar, Deepak Anand, and Amit Sethi. Classification of breast cancer histology using deep learning. In *International Conference Image Analysis and Recognition*, pages 837–844. Springer, 2018.

[39] Hongliu Cao, Simon Bernard, Laurent Heutte, and Robert Sabourin. Improve the performance of transfer learning without fine-tuning using dissimilarity-based multi-view learning for breast cancer histology images. In *International Conference Image Analysis and Recognition*, pages 779–787. Springer, 2018.

[40] Mohammed Safwan, Sai Saketh Chennamsetty, Avinash Kori, Varghese Alex, and Ganapathy Krishnamurthi. Classification of breast cancer and grading of diabetic retinopathy & macular edema using ensemble of pre-trained convolutional neural networks. *arxiv*, 2018.

[41] 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.

[42] Philipp Kainz, Michael Pfeiffer, and Martin Urschler. Semantic segmentation of colon glands with deep convolutional neural networks and total variation segmentation. *arXiv preprint arXiv:1511.06919*, 2015.

[43] Aïcha BenTaieb, Jeremy Kawahara, and Ghassan Hamarneh. Multi-loss convolutional networks for gland analysis in microscopy. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 642–645. IEEE, 2016.

[44] 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.

[45] Yan Xu, Yang Li, Mingyuan Liu, Yipei Wang, Yubo Fan, Maode Lai, and Eric I.-Chao Chang. Gland instance segmentation by deep multichannel neural networks. *arXiv:1607.04889 [cs]*, 2016.

[46] Aïcha BenTaieb and Ghassan Hamarneh. Topology aware fully convolutional networks for histology gland segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 460–468. Springer, 2016.

[47] 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.

[48] 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. *arXiv preprint arXiv:1606.00915*, 2016.

[49] Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1529–1537, 2015.

[50] 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.

[51] A. Paul and D. P. Mukherjee. Gland segmentation from histology images using informative morphological scale space. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 4121–4125, 2016.

[52] Lin Yang, Yizhe Zhang, Jianxu Chen, Siyuan Zhang, and Danny Ziyi Chen. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In *MICCAI*, 2017.

[53] 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 *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*, Lecture Notes in Computer Science, pages 408–416. Springer, Cham, 2017.

[54] Malay Singh, Emarene Mationg Kalaw, Danilo Medina Giron, Kyeong Ock Chong, Chew Lim Tan, and Hwee Kuan Lee. Gland segmentation in prostate histopathological images. *Journal of medical imaging*, 4 2:027501, 2017.

[55] Yuanpu Xie, Zizhao Zhang, Manish Sapkota, and Lin Yang. Spatial clockwork recurrent neural network for muscle perimysium segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 185–193. Springer, 2016.

[56] Sachin Mehta, Ezgi Mercan, Jamen Bartlett, Donald L. Weaver, Joann G. Elmore, and Linda G. Shapiro. Learning to segment breast biopsy whole slide images. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 663–672, 2018.

[57] Humayun Irshad, Antoine Veillard, Ludovic Roux, and Daniel Racoceanu. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. *IEEE reviews in biomedical engineering*, 7:97–114, 2014.

[58] 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. *arXiv preprint arXiv:1702.05747*, 2017.

[59] Adnan M Khan, Nasir Rajpoot, Darren Treanor, and Derek Magee. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *Biomedical Engineering, IEEE Transactions on*, 61(6):1729–1738, 2014.

[60] Andrew Janowczyk, Ajay Basavanhally, and Anant Madabhushi. Stain normalization using sparse autoencoders (stanosa): Application to digital pathology. *Computerized Medical Imaging and Graphics*, 57:50–61, 2017.

[61] Ajay Basavanhally and Anant Madabhushi. Em-based segmentation-driven color standardization of digitized histopathology. In *Proc. SPIE*, volume 8676, page 86760G. Citeseer, 2013.

[62] Chanho Jung and Changick Kim. Segmenting clustered nuclei using h-minima transform-based marker extraction and contour parameterization. *IEEE transactions on biomedical engineering*, 57(10):2600–2604, 2010.

[63] Carolina Wählby, I-M SINTORN, Fredrik Erlandsson, Gunilla Borgefors, and Ewert Bengtsson. Combining intensity, edge and shape information for 2d and 3d segmentation of cell nuclei in tissue sections. *Journal of microscopy*, 215(1):67–76, 2004.

[64] Eric Cosatto, Matt Miller, Hans Peter Graf, and John S Meyer. Grading nuclear pleomorphism on histological micrographs. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[65] Gareth Loy and Alexander Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Transactions on pattern analysis and machine intelligence*, 25(8):959–973, 2003.

[66] Mitko Veta, A Huisman, Max A Viergever, Paul J van Diest, and Josien PW Pluim. Marker-controlled watershed segmentation of nuclei in h&e stained breast cancer biopsy images. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 618–621. IEEE, 2011.

[67] Xin Qi, Fuyong Xing, David J Foran, and Lin Yang. Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set. *IEEE Transactions on Biomedical Engineering*, 59(3):754–765, 2012.

[68] Bahram Parvin, Qing Yang, Ju Han, Hang Chang, Bjorn Rydberg, and Mary Helen Barcellos-Hoff. Iterative voting for inference of structural saliency and characterization of subcellular events. *IEEE Transactions on Image Processing*, 16(3):615–623, 2007.

[69] Weidi Xie, J Alison Noble, and Andrew Zisserman. Microscopy cell counting and detection with fully convolutional regression networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pages 1–10, 2016.

[70] Saad Ullah Akram, Juho Kannala, Lauri Eklund, and Janne Heikkilä. Cell segmentation proposal network for microscopy image analysis. In *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 21–29. Springer, 2016.

[71] Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee-Wah Tsang, David Snead, Ian Cree, and Nasir Rajpoot. A spatially constrained deep learning framework for detection of epithelial tumor nuclei in cancer histology images. In *International Workshop on Patch-based Techniques in Medical Imaging*, pages 154–162. Springer, 2015.

[72] Zheng Xu and Junzhou Huang. Detecting 10,000 cells in one second. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 676–684. Springer, 2016.

[73] Jun Xu, Lei Xiang, Qingshan Liu, Hannah Gilmore, Jianzhong Wu, Jinghai Tang, and Anant Madabhushi. Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. *IEEE transactions on medical imaging*, 35(1):119–130, 2016.

[74] Florence Cloppet and Arnaud Boucher. Segmentation of overlapping/aggregating nuclei cells in biological images. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[75] Hussain Fatakdawala, Jun Xu, Ajay Basavanhally, Gyan Bhanot, Shridar Ganesan, Michael Feldman, John E Tomaszewski, and Anant Madabhushi. Expectation–maximization-driven geodesic active contour with overlap resolution (emagacor): Application to lymphocyte segmentation on breast cancer histopathology. *IEEE Transactions on Biomedical Engineering*, 57(7):1676–1689, 2010.

[76] Yousef Al-Kofahi, Wiem Lassoued, William Lee, and Badrinath Roysam. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering*, 57(4):841–852, 2010.

[77] Chanho Jung, Changick Kim, Seoung Wan Chae, and Sukjoong Oh. Unsupervised segmentation of overlapped nuclei using bayesian classification. *IEEE Transactions on Biomedical Engineering*, 57(12):2825–2832, 2010.

[78] Maria S Kulikova, Antoine Veillard, Ludovic Roux, and Daniel Racoceanu. Nuclei extraction from histopathological images using a marked point process approach. In *Medical Imaging: Image Processing*, page 831428, 2012.

[79] Antoine Veillard, Maria S Kulikova, and Daniel Racoceanu. Cell nuclei extraction from breast cancer histopathologyimages using colour, texture, scale and shape information. *Diagnostic Pathology*, 8(1):S5, 2013.

[80] Hui Kong, Metin Gurcan, and Kamel Belkacem-Boussaid. Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting. *IEEE transactions on medical imaging*, 30(9):1661–1677, 2011.

[81] 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.

[82] Youyi Song, Ee-Leng Tan, Xudong Jiang, Jie-Zhi Cheng, Dong Ni, Siping Chen, Baiying Lei, and Tianfu Wang. Accurate cervical cell segmentation from overlapping clumps in pap smear images. *IEEE transactions on medical imaging*, 36(1):288–300, 2017.

[83] Fuyong Xing, Yuanpu Xie, and Lin Yang. An automatic learning-based framework for robust nucleus segmentation. *IEEE transactions on medical imaging*, 35(2):550–566, 2016.

[84] Hai Su, Fujun Liu, Yuanpu Xie, Fuyong Xing, Sreenivasan Meyyappan, and Lin Yang. Region segmentation in histopathological breast cancer images using deep convolutional neural network. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 55–58. IEEE, 2015.

[85] M Murat Dundar, Sunil Badve, Gokhan Bilgin, Vikas Raykar, Rohit Jain, Olcay Sertel, and Metin N Gurcan. Computerized classification of intraductal breast lesions using histopathological images. *IEEE Transactions on Biomedical Engineering*, 58(7):1977–1984, 2011.

[86] Ludovic Roux, Daniel Racoceanu, Nicolas Loménie, Maria Kulikova, Humayun Irshad, Jacques Klossa, Frédérique Capron, Catherine Genestie, Gilles Le Naour, Metin N Gurcan, et al. Mitosis detection in breast cancer histological images an icpr 2012 contest. *Journal of pathology informatics*, 4(1):8, 2013.

[87] Christopher D Malon and Eric Cosatto. Classification of mitotic figures with convolutional neural networks and seeded blob features. *Journal of pathology informatics*, 4, 2013.

[88] Anat Shkolyar, Amit Gefen, Dafna Benayahu, and Hayit Greenspan. Automatic detection of cell divisions (mitosis) in live-imaging microscopy images using convolutional neural networks. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 743–746. IEEE, 2015.

[89] Jiawen Yao, Sheng Wang, Xinliang Zhu, and Junzhou Huang. Imaging biomarker discovery for lung cancer survival prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 649–657. Springer, 2016.

[90] 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 transactions on medical imaging*, 35(5):1196–1206, 2016.

[91] 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.

[92] Joseph G. Jacobs, Gabriel J. Brostow, Alex Freeman, Daniel C. Alexander, and Eleftheria Panagiotaki. Detecting and classifying nuclei on a budget. In *CVII-STENT/LABELS@MICCAI*, 2017.

[93] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[94] Pedro Costa, Adrian Galdran, Maria Ines Meyer, Meindert Niemeijer, Michael Abràmoff, Ana Maria Mendonça, and Aurélio Campilho. End-to-end adversarial retinal image synthesis. *IEEE transactions on medical imaging*, 8, 2017.

[95] Lei Bi, Jinman Kim, Ashnil Kumar, Dagan Feng, and Michael Fulham. Synthesis of positron emission tomography (pet) images via multi-channel generative adversarial networks (gans). In *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, pages 43–51. Springer, 2017.

[96] Neslihan Bayramoglu, Mika Kaakinen, Lauri Eklund, and Janne Heikkilä. Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks. *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 64–71, 2017.

[97] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind, Wenda Wang, and Russell Webb. Learning from simulated and unsupervised images through adversarial training. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2242–2251, 2017.

[98] Francesco Calimeri, Aldo Marzullo, Claudio Stamile, and Giorgio Terracina. Biomedical data augmentation using generative adversarial neural networks. In *International Conference on Artificial Neural Networks*, pages 626–634. Springer, 2017.

[99] Jian Zhao, Lin Xiong, Panasonic Karlekar Jayashree, Jianshu Li, Fang Zhao, Zhecan Wang, Panasonic Sugiri Pranata, Panasonic Shengmei Shen, Shuicheng Yan, and Jiashi Feng. Dual-agent gans for photorealistic and identity preserving profile face synthesis. In *Advances in Neural Information Processing Systems*, pages 66–76, 2017.

[100] Anton Osokin, Anatole Chessel, Rafael E Carazo Salas, and Federico Vaggi. Gans for biological image synthesis. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2252–2261. IEEE, 2017.

[101] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, volume 2, page 4, 2017.

[102] Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Advances in Neural Information Processing Systems*, pages 82–90, 2016.

[103] Michael Mathieu, Camille Couprie, and Yann LeCun. Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440*, 2015.

[104] Anh Nguyen, Jeff Clune, Yoshua Bengio, Alexey Dosovitskiy, and Jason Yosinski. Plug & play generative networks: Conditional iterative generation of images in latent space. In *CVPR*, volume 2, page 7, 2017.

[105] Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pages 1486–1494, 2015.

[106] Daniel Jiwoong Im, Chris Dongjoo Kim, Hui Jiang, and Roland Memisevic. Generating images with recurrent adversarial networks. *arXiv preprint arXiv:1602.05110*, 2016.

[107] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8798–8807, 2018.

[108] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. In *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science, pages 649–666. Springer, Cham, 2016.

[109] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *arXiv preprint arXiv:1809.07294*, 2018.

[110] John T Guibas, Tejpal S Virdi, and Peter S Li. Synthetic medical images from dual generative adversarial networks. *arXiv preprint arXiv:1709.01872*, 2017.

[111] Peter Goldsborough, Nick Pawlowski, Juan C Caicedo, Shantanu Singh, and Anne Carpenter. Cytogan: Generative modeling of cell images. *bioRxiv*, page 227645, 2017.

[112] Faisal Mahmood, Daniel Borders, Richard Chen, Gregory N McKay, Kevan J Salimian, Alexander Baras, and Nicholas J Durr. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *arXiv preprint arXiv:1810.00236*, 2018.

[113] Lei Bi, Dagan Feng, and Jinman Kim. Dual-path adversarial learning for fully convolutional network (fcn)-based medical image segmentation. *The Visual Computer*, pages 1–10, 2018.

[114] Caglar Senaras, Muhammad Khalid Khan Niazi, Berkman Sahiner, Michael P Pennell, Gary Tozbikian, Gerard Lozanski, and Metin N Gurcan. Optimized generation of high-resolution phantom images using cgan: Application to quantification of ki67 breast cancer images. *PloS one*, 13(5):e0196846, 2018.

[115] Du Wang, Chaochen Gu, Kaijie Wu, and Xinping Guan. Adversarial neural networks for basal membrane segmentation of microinvasive cervix carcinoma in histopathology images. In *Machine Learning and Cybernetics (ICMLC), 2017 International Conference on*, volume 2, pages 385–389. IEEE, 2017.

[116] 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.

[117] Bo Hu, Ye Tang, Eric I Chang, Yubo Fan, Maode Lai, Yan Xu, et al. Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks. *arXiv preprint arXiv:1711.11317*, 2017.

[118] Farhad Ghazvinian Zanjani, Svitlana Zinger, Babak Ehteshami Bejnordi, Jeroen AWM van der Laak, and Peter HN de With. Stain normalization of histopathology images using generative adversarial networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 573–577. IEEE, 2018.

[119] Aman Rana, Gregory Yauney, Alarice Lowe, and Pratik Shah. Computational histological staining and destaining of prostate core biopsy rgb images with generative adversarial neural networks. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 828–834. IEEE, 2018.

[120] Oren Rippel and Lubomir Bourdev. Real-time adaptive image compression. *arXiv preprint arXiv:1705.05823*, 2017.

[121] Yair Rivenson, Hongda Wang, Zhensong Wei, Yibo Zhang, Harun Gunaydin, and Aydogan Ozcan. Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue. *arXiv:1803.11293 [physics]*, 2018.

[122] Nicholas Trahearn, David Epstein, Ian A. Cree, David R. J. Snead, and Nasir M. Rajpoot. Hyper-stain inspector: A framework for robust registration and localised co-expression analysis of multiple whole-slide images of serial histology sections. In *Scientific Reports*, 2017.

[123] Akif Burak Tosun, Luong Nguyen, Nathan Ong, Olga Navolotskaia, Gloria Carter, Jeffrey L. Fine, D. Lansing Taylor, and S. Chakra Chennubhotla. Histological detection of high-risk benign breast lesions from whole slide images. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*, pages 144–152. Springer, Cham, 2017.

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

[125] V Wiesmann, D Franz, C Held, C Münzenmayer, R Palmisano, and T Wittenberg. Review of free software tools for image analysis of fluorescence cell micrographs. *Journal of microscopy*, 257(1):39–53, 2015.

[126] Peter Bajcsy, Antonio Cardone, Joe Chalfoun, Michael Halter, Derek Juba, Marcin Kociolek, Michael Majurski, Adele Peskin, Carl Simon, Mylene Simon, et al. Survey statistics of automated segmentations applied to optical imaging of mammalian cells. *BMC bioinformatics*, 16(1):330, 2015.

[127] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7):671–675, 2012.

[128] Michael R Lamprecht, David M Sabatini, Anne E Carpenter, et al. Cellprofiler™: free, versatile software for automated biological image analysis. *Biotechniques*, 42(1):71, 2007.

[129] R Marée, L Rollus, B Stévens, R Hoyoux, G Louppe, R Vandaele, J-M Begon, P Kainz, P Geurts, and L Wehenkel. Cytomine: An open-source software for collaborative analysis of whole-slide images. *Diagnostic Pathology*, 1(8), 2016.

[130] Raphaël Marée, Loïc Rollus, Benjamin Stévens, Renaud Hoyoux, Gilles Louppe, Rémy Vandaele, Jean-Michel Begon, Philipp Kainz, Pierre Geurts, and Louis Wehenkel. Collaborative analysis of multi-gigapixel imaging data using cytomine. *Bioinformatics*, 32(9):1395–1401, 2016.

[131] 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.

[132] Thouis R Jones, In Han Kang, Douglas B Wheeler, Robert A Lindquist, Adam Papallo, David M Sabatini, Polina Golland, and Anne E Carpenter. Cellprofiler analyst: data exploration and analysis software for complex image-based screens. *BMC bioinformatics*, 9(1):482, 2008.

[133] David Dao, Adam N Fraser, Jane Hung, Vebjorn Ljosa, Shantanu Singh, and Anne E Carpenter. Cellprofiler analyst: interactive data exploration, analysis and classification of large biological image sets. *Bioinformatics*, 32(20):3210–3212, 2016.

[134] Ilya G Goldberg, Chris Allan, Jean-Marie Burel, Doug Creager, Andrea Falconi, Harry Hochheiser, Josiah Johnston, Jeff Mellen, Peter K Sorger, and Jason R Swedlow. The open microscopy environment (ome) data model and xml file: open tools for informatics and quantitative analysis in biological imaging. *Genome biology*, 6(5):R47, 2005.

[135] 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.

[136] Ulysse Rubens, Renaud Hoyoux, Laurent Vanosmael, Mehdy Ouras, Maxime Tasset, Christopher Hamilton, Rémi Longuespée, and Raphaël Marée. Cytomine: Toward an open and collaborative software platform for digital pathology bridged to molecular investigations. *PROTEOMICS–Clinical Applications*, page 1800057, 2018.

[137] Song-Chun Zhu, Cheng-En Guo, Yizhou Wang, and Zijian Xu. What are textons? *International Journal of Computer Vision*, 62(1-2):121–143, 2005.

[138] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[139] 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.

[140] 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.

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

[142] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[143] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[144] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

[145] Shaimaa Al-Janabi, A. Huisman, and Paul J van Diest. Digital pathology: current status and future perspectives. *Histopathology*, 61 1:1–9, 2012.

[146] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(Dec):3371–3408, 2010.

[147] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[148] Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

[149] Jonathan Masci, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. *Artificial Neural Networks and Machine Learning–ICANN 2011*, pages 52–59, 2011.

[150] Chih-Jen Lin. Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 19(10):2756–2779, 2007.

[151] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1):89–97, 2004.

[152] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *International journal of computer vision*, 22(1):61–79, 1997.

[153] Pablo Marquez-Neila, Luis Baumela, and Luis Alvarez. A morphological approach to curvature-based evolution of curves and surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):2–17, 2014.

[154] 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.

[155] Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1175–1183. IEEE, 2017.

[156] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[157] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[158] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[159] Dumitru Erhan, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov. Scalable object detection using deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2147–2154, 2014.

[160] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *arXiv preprint arXiv:1708.02002*, 2017.

[161] Digital pathology: Segmentation of nuclei in images. *MICCAI CPM - Competition*, 2017.

[162] Mulit-organ nuclei segmentation challenge. *MICCAI 2018 Competition*, 2018.

[163] 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.

[164] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[165] Mark Everingham, Luc Gool, Christopher K. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vision*, 88(2):303–338, June 2010.

[166] Scott Doyle, Shannon Agner, Anant Madabhushi, Michael Feldman, and John Tomaszewski. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 496–499. IEEE, 2008.

[167] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6997–7005, 2017.

[168] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv:1411.1784 [cs, stat]*, abs/1411.1784, 2014.

[169] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint*, 2017.

[170] Ming-Yu Liu and Oncel Tuzel. Coupled generative adversarial networks. In *Advances in neural information processing systems*, pages 469–477, 2016.

[171] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017.

[172] Li Liu, Jie Chen, Paul Fieguth, Guoying Zhao, Rama Chellappa, and Matti Pietikainen. A survey of recent advances in texture representation. *arXiv preprint arXiv:1801.10324*, 3, 2018.

[173] Michal Haindl and Stanislav Mikes. Texture segmentation benchmark. In *2008 19th International Conference on Pattern Recognition*, pages 1–4. IEEE, 2008.

[174] Yaniv Taigman, Adam Polyak, and Lior Wolf. Unsupervised cross-domain image generation. *CoRR*, abs/1611.02200, 2017.

[175] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2642–2651. JMLR. org, 2017.

[176] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, volume 2, pages 1398–1402. Ieee, 2003.

[177] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016.