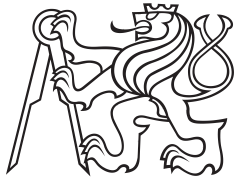


Master Thesis



**Czech
Technical
University
in Prague**

F3

**Faculty of Electrical Engineering
Department of Computer Science**

Classification of tumor type from histopathological images

Jan Kúdelka

**Supervisor: prof. Dr. Ing. Jan Kybic
Field of study: Machine Learning
Subfield: Deep Learning
December 2020**

I. Personal and study details

Student's name: **Kúdelka Jan** Personal ID number: **456945**
Faculty / Institute: **Faculty of Electrical Engineering**
Department / Institute: **Department of Computer Science**
Study program: **Open Informatics**
Specialisation: **Artificial Intelligence**

II. Master's thesis details

Master's thesis title in English:

Classification of tumor type from histopathological images

Master's thesis title in Czech:

Klasifikace typu nádoru z histopatologických obrazů

Guidelines:

Given the histological microscopy images from the PETACC3 trial:
- Get acquainted with the data and the related software. Perform literature survey on CNN-based image segmentation and classification methods.
- Design, implement and evaluate a method for segmenting the images into normal tissue, tumor tissue, and background.
- Design, implement and evaluate a method for classifying the images according to tumor type, i.e. whether it is mucinous, serrated, or Crohn-like.

Bibliography / sources:

- [1] Popovici V, Budinska E. et al. "Identification of a Poor-Prognosis BRAF-Mutant-Like Population of Patients With Colon Cancer", Journal of Clinical Oncology, pp.1288-1295, 2012,
- [2] Budinska E., Popovici V. et al. "Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer", Journal of Pathology, pp. 63-76, 2013
- [3] Vesal et al: Classification of breast cancer histology images using transfer learnin. Imagee Analysis and Reecognition, pp.812-819, 2018
- [4] Xu et al: Large Scale Tissue Histopathology Image Classification, Segmentation, and Visualization via Deep Convolutional Activation Features. BMC Bioinformatics. pp. 281, 2017
- [5] Bardou et al, Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks. IEEE Access, p.24680, 2018
- [6] Shaban et al: Context-Aware Convolutional Neural Network for Grading of Colorectal Cancer Histology Images. IEEE Trans. Medical Imaging, 2020

Name and workplace of master's thesis supervisor:

prof. Dr. Ing. Jan Kybic, Biomedical imaging algorithms, FEE

Name and workplace of second master's thesis supervisor or consultant:

Date of master's thesis assignment: **29.07.2020** Deadline for master's thesis submission: **05.01.2021**

Assignment valid until: **19.02.2022**

prof. Dr. Ing. Jan Kybic
Supervisor's signature

Head of department's signature

prof. Mgr. Petr Páta, Ph.D.
Dean's signature

III. Assignment receipt

The student acknowledges that the master's thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the master's thesis, the author must state the names of consultants and include a list of references.

Date of assignment receipt

Student's signature

Acknowledgements

I would like to thank my supervisor prof. Jan Kybic for guiding me during my work on my master thesis. I would also like to thank my parents Jana and Ivo for their unwavering support during my entire studies.

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

Prague, 23. December 2020

Abstract

In recent years, machine learning has been used increasingly more often in most areas of science and engineering. One such area is the analysis of data from the medical environment. This master thesis is concerned with the application of machine learning on the analysis of histopathological images. The main challenge of the analysis of such images is their size. Each image has a size of up to a few gigabytes. For this reason, it is not possible to use conventional methods of machine learning for their analysis. In this work, we propose a robust classifier, which can detect tumours in the images and classify them, despite the challenges that come with it. To reach this goal, data preprocessing as well as deep learning and multiple instance learning methods are used. The implementation of used methods has been verified on known datasets.

Keywords: Digital Pathology, Deep Learning, Machine learning, Whole Slide Images

Supervisor: prof. Dr. Ing. Jan Kybic

Abstrakt

V poslední době stoupá frekvence použití strojového učení v celé řadě oblastí vědy a techniky. Jednou z těchto oblastí je i analýza dat z medicinského prostředí. Tato diplomová práce se zabývá využitím strojového učení k analýze histopatologických snímků. Hlavní výzvou při zpracování digitálních histopatologických snímků je jejich velikost. Každý snímek dosahuje velikosti až několik gigabytů. Z tohoto důvodu není možné použít k jejich analýze konvenční metody strojového učení. Cílem práce je navrhnout a naimplementovat robustní klasifikátor, který zvládne ve snímcích detekovat a klasifikovat nádory, navzdory výzvám, které jsou s tím spjaté. K dosažení cíle je využíváno předzpracování dat, prvky hlubokého učení a metoda "multiple instance learning". Implementace použitých metod byly ověřeny na známých datasetech.

Klíčová slova: Digitální patologie, Hluboké učení, Strojové učení, Histopatologické snímky

Překlad názvu: Klasifikace typu nádoru z histopatologických obrazů

Contents

1 Introduction	1	2.3.4 Fully Convolutional Networks	12
1.1 Task Definition	2	2.3.5 Convolutional Models with Graphical Models	13
1.2 Thesis Structure	3	2.3.6 Deep Encoder-Decoder Models	14
2 Literature review	5	3 Essential Algorithms and Structures	17
2.1 Digital Imaging in Medicine	5	3.1 Convolutional Neural Networks	17
2.2 Aspects of Whole Slide Imaging	6	3.1.1 Convolutional Layer	18
2.2.1 Data Availability	6	3.1.2 Pooling Layer	18
2.2.2 Data Variability and Artefacts	7	3.1.3 Fully-connected Layer	20
2.2.3 Large Data Size	7	3.1.4 Activation Functions	20
2.2.4 Low Signal to Noise Ratio	8	3.1.5 Batch-normalisation	21
2.2.5 Low Interpretability of DL Methods	9	3.1.6 Dropout	22
2.3 Deep Learning Methods for Image Analysis	9	3.1.7 Inception Architecture	22
2.3.1 History of Convolutional Neural Networks	10	3.2 Multiple Instance Learning Algorithms	23
2.3.2 CNN Innovations	11	3.2.1 mi-SVM	24
2.3.3 CNN-based Image Segmentation Methods	11	3.2.2 Ratio-constrained Multiple Instance Markov Network (RMIMN)	24

4 Methodology	29	5.1.1 Results	42
4.1 PETACC3 Dataset	29	5.1.2 Example Segmentations	42
4.2 WSI Segmentation Methodology	30	5.2 Experiment 2 - WSI Segmentation into Tumorous and Normal Tissue, CNN Trained on All Images	43
4.3 Segmentation into Tissue, Background and Pen Marks	32	5.2.1 Results	44
4.4 Patch Extraction	33	5.3 Experiment 3 - WSI Classification According to Tumour Type - Direct Method	44
4.5 Model Architecture and Training	33	5.3.1 Results	44
4.5.1 Input Transformations	34	5.4 Experiment 4 - mi-SVM trained on MUSK and MUSK2 datasets	45
4.5.2 Network Parameters and Training	35	5.4.1 Results	45
4.6 WSI Classification Methodology	35	5.5 Experiment 5 - WSI Classification According to Tumor Type - mi-SVM	45
4.7 MIL Approach	36	5.5.1 Results	46
4.7.1 Patch Description	36	5.6 Experiment 6 - WSI Classification According to Tumor Type - RMIMN	46
4.7.2 Bag Classification	37	5.6.1 Results	46
4.7.3 Schema of the MIL Approach	38		
4.8 Direct Approach	38		
5 Experiments	41	6 Conclusions and Future Work	49
5.1 Experiment 1 - WSI Segmentation into Tumorous and Normal Tissue, CNN Trained on Pure-case Images	41	6.1 Conclusions	49
		6.2 Future Work	50

A Attachments 51

B Bibliography 53

Figures

1.1 Example WSI from the PETACC3 dataset.	2
2.1 An example of artefact detection in WSI. Source:[12]	7
2.2 An example of patch extraction. Images were created using the PETACC3 dataset [72]. A WSI from the dataset on the left, extracted patches on the right.	8
2.3 LeNet-5 architecture Source: [49]	10
2.4 CNN architectures. Adapted from: [39]	11
2.5 FCN architecture. Source: [54] .	13
2.6 CNN + CRF architecture. Source: [23]	13
2.7 Encoder-Decoder architecture using deconvolution. Source: [58] .	14
2.8 U-net architecture. Source: [60]	15
3.1 Example convolutional layer operation. One receptive field and its corresponding output value have been marked blue.	19
3.2 Example max-pooling layer operation. One receptive field and its corresponding output value have been marked blue.	19
3.3 Standard CNN activation functions.	21
3.4 A visualisation of Dropout.	22
3.5 Schema of the Inception block. Adapted from [67].	23
3.6 Pseudocode of the mi-SVM algorithm. Source: [15].	24
3.7 Schema of the proposed RMIMN model. Source: [30].	25
4.1 Example image from the PETACC3 dataset. Marked pure-case with labels: mucinous=1, serrated=2, Crohn-like=2.	31
4.2 Schema of the entire proposed WSI segmentation method.	32
4.3 Example Background/Tissue segmentation. Original down-sampled image on the left, segmentation mask on the right. Tissue in light grey, marker lines in darker grey, background in dark grey.	33
4.4 Example extracted patches from the PETACC 3 dataset. Patches were labelled TU and NO respectively. Both patches have dimensions of 1024x1024 pxs.	34

4.5 Schema of the proposed MIL approach. 39

5.1 Example Normal/Tumorous tissue segmentation. Tissue patches classified as normal tissue are marked with a green border. Patches classified as tumorous are marked with a red border. The image is from the pure-case set. 42

5.2 Example Normal/Tumorous tissue segmentation. Tissue patches classified as normal tissue are marked with a green border. Patches classified as tumorous are marked with a red border. The image is from the pure-case set. 43

Tables



Chapter 1

Introduction

Gigapixel image analysis is the task of analysing images with more than a billion pixels. This is by no means a trivial task, mainly because of the sheer size of each image. Such a large size prevents the use of conventional machine learning methods, such as the direct use of convolutional neural networks, which have been the leading instrument in automatic segmentation and classification of images in recent years [56, 29].

One type of gigapixel images is the Whole-slide images (or WSIs). A WSI is a high-resolution image of a whole microscope slide. This format is often used in histopathology and many other areas of medicine such as neuroanatomy, proteomics (the large-scale study of proteins), connectomics (the study of maps of connections in an organism's neural system) and genomics (the study of the function, structure and editing of genes) [11]. The analysis of whole-slide images is one of the leading challenges in computer vision in medicine. Competitions are held yearly to evaluate existing and new algorithms for their classification and segmentation [3]. The state-of-the-art algorithms that are devised in these competitions have the potential to make predictions about these images faster and more precisely.

The PETACC3 trial was a phase III trial (a trial where the new drug is compared to the standard-of-care drug) with random assignment of treatment in multiple medical centers to test the effectivity of the addition of irinotecan to other forms of medication for patients with stage 3 colorectal cancer [3]. In this thesis, we are working with microscopy images extracted during this trial.

1.1 Task Definition

Given the histological microscopy images from the PETACC3 trial, the goal of this thesis is defined as to:

- Get acquainted with the data and the related software. Perform literature survey on CNN-based image segmentation and classification methods.
- Design, implement and evaluate a method for segmenting the images into normal tissue, tumour tissue and background.
- Design, implement and evaluate a method for classifying the images according to tumour type, i.e. whether it is mucinous, serrated or Crohn-like.

In order to accomplish this task, we propose an extensive survey of state-of-the-art methods of image segmentation and WSI analysis. Then, we adopt some of these methods and implement them. Finally, experiments are done to evaluate the performance of these methods on the PETACC3 dataset. An example image from the dataset can be seen in Figure 1.1.



Figure 1.1: Example WSI from the PETACC3 dataset.

■ 1.2 Thesis Structure

Chapter 2: Literature Review

This chapter contains an overview of modern image segmentation and classification methods, focusing on convolutional neural networks. The aspects of Whole Slide Imaging are also reviewed in this chapter.

Chapter 3: Essential Algorithms and Structures

This chapter contains descriptions of the algorithms and structures used in this thesis. A more in-depth approach is taken to acquaint the reader with the methods used in this thesis.

Chapter 4: Methodology of WSI Segmentation into Tumour and Normal Tissue

The proposed methods for WSI segmentation into tumorous and normal tissue are discussed in this chapter.

Chapter 5: Methodology

The proposed methods for WSI segmentation as well as the proposed methods of classifying the WSIs according to tumour type is presented in this chapter.

Chapter 5: Experiments

We propose several experiments to evaluate the performance of our proposed method in this chapter

Chapter 6: Conclusion and Future Work

Conclusions are made about the results of our experiments. Future work on this subject is also discussed.



Chapter 2

Literature review

Whole-slide imaging is the process of digitising an entire microscopy slide into a single image file. With the emerging of this method using the appropriate scanners, digital images have seen a substantial increase in use in pathology and other areas of medicine and medical research [79]. Many studies have shown the feasibility of using these images in practical use [14, 78]. The main benefits of digital images in pathology are their remote accessibility, the possibility of easy, long-term storage, and, with the development of machine learning algorithms for image processing, their automatic processing and analysis.



2.1 Digital Imaging in Medicine

The practical use of digital imaging comes with many challenges. First of all, scanners that are used for image digitisation have to be approved by regulatory institutes in each respective country. For example, in the United States of America, only one such scanner has been approved to date for primary diagnosis [2]. Even with approved scanners, introducing digitisation is always very expensive. Each scanner goes for as much as hundreds of thousands of US dollars [45]. Further costs come from training the staff to operate these scanners and obtaining new hardware for storing and processing these images. Despite these challenges, most modern medical facilities are using some digitisation. Often, the digital images are used for second opinions, long-term storage, or teaching purposes [45, 79].

■ 2.2 Aspects of Whole Slide Imaging

In the last decade, WSI has shown promise as a base for the analysis of pathological images by deep learning algorithms. Deep learning has stood as the best means of automatic image analysis for some time since the introduction of convolutional neural networks at the end of the last century [48, 49, 29]. Applying such a powerful tool to WSIs has the potential to make very fast, precise and robust predictions about these digital images. However, to successfully design and implement such an algorithm, we must first understand what features, challenges and anomalies can occur when dealing with digital histological images. Many such challenges occur, especially when aiming for clinical relevance [40, 52, 69, 70]. Essential challenges and their solutions or workarounds are mentioned in this section.

■ 2.2.1 Data Availability

When using DL algorithms, a sufficient amount of training samples is necessary for achieving high accuracy [85, 50, 35]. Ideally, these samples should also be well-annotated by experts. However, labelling histopathological images is a very long, costly and tiresome process, especially when dealing with more complex classes (e.g. mucinous), as opposed to binary classification (e.g. existence of a tumour anywhere in the image) [40, 70]. Furthermore, medical data is often under restrictions due to its sensitive and private nature[10].

In spite of these aspects, labelled rich datasets have started emerging in the last couple of years. To name a few:

- CAMELYON: 1399 H&E-stained (Hematoxylin and eosin) sentinel lymph node sections of breast cancer patients [51]
- BACH: 400 H&E stained breast histology microscopy and whole-slide images used for the ICIAR 18 challenge [1]
- BreCaHAD: a dataset for breast cancer histopathological annotation and diagnosis [13]

2.2.2 Data Variability and Artefacts

There is a high level of variability when it comes to WSI datasets. This is the case for two reasons. Firstly, a WSI dataset can come from multiple sources at once. For this reason, each image might have been obtained using different methods specific to the source (e.g. different staining methods, different scanning device). Secondly, artefacts are often introduced when the images are being processed. This includes, but is not limited to uneven illumination and focus, tissue tears and fold, and pen marks[26, 70]. This variability needs to be addressed when dealing with such datasets, and proper generalisation must be ensured.

Kothari, Phan and Wang in [42] suggest the use of saturation and intensity values to classify each pixel of the image and detect tissue folds in this manner. They show a significant increase in the performance of cancer detection models after applying their method to WSIs.

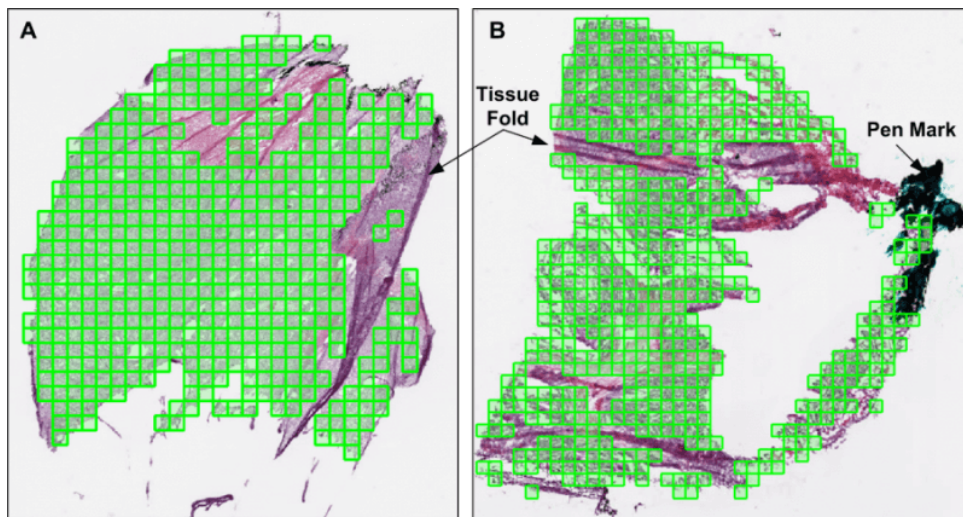


Figure 2.1: An example of artefact detection in WSI. Source:[12]

2.2.3 Large Data Size

With a typical resolution of 100 000x100 000 pixels and a typical size of a few gigabytes, WSIs become difficult to analyse. This difficulty applies to both classification and segmentation. Such a size prevents the use of conventional DL methods (e.g. the direct use of convolutional neural networks), due to hardware limitations. Instead, some workarounds need to be implemented[26].

The most common way of dealing with this issue is the patch extraction method. Many researchers have used this method to achieve state-of-the-art results [16, 18, 19]. By using this method, the image is split into smaller, square patches (typically 200-1000 pixels in each dimension).

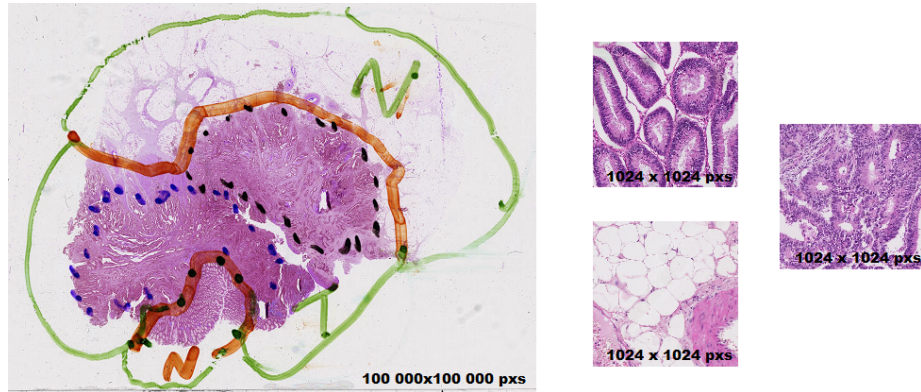


Figure 2.2: An example of patch extraction. Images were created using the PETACC3 dataset [72]. A WSI from the dataset on the left, extracted patches on the right.

2.2.4 Low Signal to Noise Ratio

WSIs often have a low signal to noise ratio. Only a small part of the image is linked to the image label (e.g. malignant cells), while the rest of the image is irrelevant (e.g. background, healthy cells). The image's spatial distribution can also be important but can be lost when making simplifying assumptions about the task [26, 69]. This limits the use of the patch method, described in the previous section, and assumptions have to be made about the images.

One such assumption is that each extracted patch shares its label with the image. This leads to a naive solution, where many labelled small images are classified with an implemented classifier, and then the whole WSI is classified, using some prediction rule [69]. However, for this assumption to be feasible, the work of an expert is required, who needs to provide pixel-level annotations to training data. Methods using these strongly annotated patches are known to achieve an impressive prediction accuracy [74, 64]. Very often, however, we have to work with weakly annotated data, where the label is provided only on the image level. In this case, we can not make the same assumption without the loss of performance.

In the weakly annotated case, researchers have made a different assumption. They assumed that although the labelled aspect of the image is not recognisable in all of its extracted patches, it is still recognisable in some of

them. This leads to solutions in the multiple instance learning algorithms, which have also shown impressive results [77, 71].

The patch extraction method has been proven to be a very powerful tool in the WSI analysis. Nevertheless, spatial information about the image is always lost in the process. For this reason, other means of dimensionality reduction are being researched. In the scope of pathology, this is especially important for the detection of metastasis, which is usually not detectable using the patch extraction method [40]. Tellez et al., 2019 [69] suggest the use of neural image compression, which is the technique of mapping WIS to a higher-level latent space. The researchers in [41] propose the use of their novel network Spatio-Net, which uses a CNN to compress each patch and a 2D-Long-Short Term Memory network to classify all the compressed patches. Li and Ping, 2018 [74] make use of a neural conditional random field to take advantage of the spatial information in WSIs.

2.2.5 Low Interpretability of DL Methods

Finally, it is not easy to extract clinically relevant results from DL methods. Due to their low interpretability, deep neural networks are often treated as black boxes [21]. In other words, DL methods might provide seemingly very accurate predictions and features; however, interpreting them can very often be near impossible. For this reason, their application in the medical field might be difficult to employ completely and independently, as providing reasoning behind decisions is often required in the medical field [40, 70]. Methods of interpreting artificial neural network decisions are being researched (e.g. [57, 62, 82]), however, their applicability in medical decisions is still unclear [21].

2.3 Deep Learning Methods for Image Analysis

With the introduction of convolutional neural networks, deep learning methods have taken over as state of the art in image classification, and analysis [56]. What follows in this section is a brief history of CNNs, a summary of modern innovations, and a short survey of methods of their use in image segmentation.

2.3.1 History of Convolutional Neural Networks

One of the first introductions of a convolutional neural network was done in the late 1980s by Yann Lecun et al. [48], using backpropagation for recognition of handwritten characters. In 1998, Lecun et al. showed that their CNN LeNet-5 outperformed other techniques of handwritten character recognition [49]. However, CNNs require scaling, if they are to be applied to high-resolution images, which was not possible with the hardware at the time.

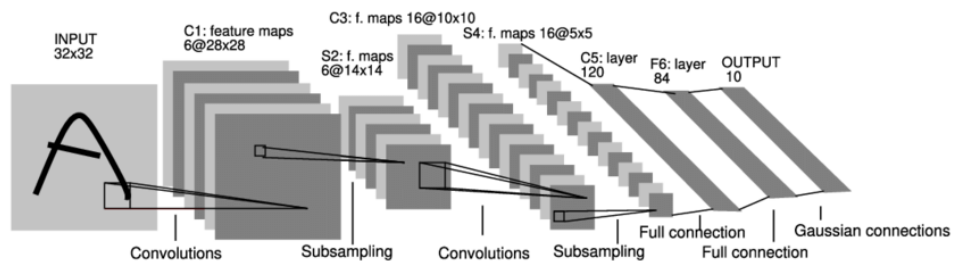


Figure 2.3: LeNet-5 architecture Source: [49]

The popularity of convolutional neural networks rose significantly with their reimplementations for Graphics Processing Units (GPUs) by Chellapilla et al. in 2006 [22]. The authors also introduced more efficient methods of the implementation of convolution, using matrix multiplications. The combination of this and the use of a GPU decreased the learning wall clock time 3-4 times. It was also predicted that this could increase significantly with larger networks.

With the establishment of the immense ImageNet database [25], and the corresponding annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [6], where new algorithms for object detection and image classification are evaluated, the superiority of CNNs became clear. In 2012, AlexNet, a deep convolutional neural network architecture designed by Alex Krizhevsky [43], won the competition, beating its runner ups by more than 10% on the top-5 error [4]. Two years later, during the ILSVRC 2014, GoogLeNet [67] and its close runner up VGG-16 [63] achieved near human-level accuracy of 6.66% and 7.32% respectively [5] on the top-5 error.

Finally, at ILSVRC 2015, ResNet was introduced. This deep convolutional neural network architecture had 152 layers (as opposed to GoogLeNet and VGG-16 with 19 and 16 respectively [67, 63]), and contained skip connections to reduce spatial complexity, with only 25.5 million parameters, roughly five

times less than VGG-16 [33]. This network achieved a classification error of only 3.57%, beating human experts for the first time [7].

Till this day, VGG-16 and GoogleNet (and its newer iterations InceptionV2 and InceptionV3 [68]), remain the most used deep convolutional neural network architectures, due to their simple design with a high level of performance.

2.3.2 CNN Innovations

In the last couple of years, many new innovations in convolutional neural networks have been popping up, which make clever use of spatial properties, data flows and residual links (a.k.a. skip connections). What follows is a table summarising these innovative network architectures, adopted from Khan et al. 2020 [39], including previously mentioned CNN architectures for comparison.

Architecture Name	Year	Main contribution	Ref
LeNet	1998	First popular CNN architecture	[49]
AlexNet	2012	Deeper and wider than the LeNet	[43]
ZfNet	2014	Visualisation of intermediate layers	[83]
VGG	2014	Homogenous topology, Small kernels	[63]
GoogLeNet	2015	Split transform and merge idea	[67]
InceptionV3	2015	Handles representational bottleneck	[68]
InceptionV4	2016	Uses asymmetric filters	[66]
Inception-ResNet	2016	Split transform and merge + res. links	[66]
ResNet	2016	Identity mapping based skip connections	[33]
DelugeNet	2016	Cross layer information flow	[44]
FractalNet	2016	Multi-path architecture without residuals	[47]
WideResNet	2016	Increased width, decreased depth	[81]
Xception	2017	Depth wise conv followed by point wise conv	[24]
DenseNet	2017	Cross-layer information flow	[37]
PolyNet	2017	Structural diversity, generalised residual units	[84]
PyramidalNet	2017	Gradual increase in width per unit	[32]

Figure 2.4: CNN architectures. Adapted from: [39]

2.3.3 CNN-based Image Segmentation Methods

With the establishment of CNNs as state of the art in image classification, researchers have studied ways of their application in image segmentation. Many such novel methods have been derived [56]. Some of these methods are described in the next sections. Specifically, we talk about:

- Fully Convolutional Networks
- Convolutional Models With Graphical Models
- Deep Encoder-Decoder Models
- U-net and V-net architectures

■ 2.3.4 Fully Convolutional Networks

One of the architectures for image segmentation that was derived from the CNNs are the Fully convolutional networks (FCN). This architecture is used for semantic segmentation, which means the output of the network is the same size as its input, mapping each pixel to a single class. The FCN architecture was first proposed by Long et al. in 2015 [54], who implemented FCNs by modifying existing CNN architectures, namely the aforementioned Alexnet, GoogLeNet and VGG-16 architectures.

As the name implies, Fully convolutional networks do not have any fully connected layers, and therefore only consist of convolutional and pooling layers. Such a network must, therefore, also have a spatial output. When modifying existing classifiers, the author first uses a process called convolutionalisation, where they replace all fully connected layers by convolutional layers with a kernel size equal to the size of the entire input. An output layer is then added, in the form of a convolutional layer with a kernel size of 1×1 and channel size equal to the number of classes + 1 for the background which represents scores for each class. Finally, a backwards convolutional layer can be added to upsample the output back to the original image size. A backwards convolutional layer is a convolutional layer that has had its backwards and forwards message switched, resulting in convolution with a stride of $1/f$. The whole network is then trained using a standard Stochastic Gradient Descent algorithm.

Fully convolutional neural networks were shown to achieve great performance for image segmentation, achieving a relative 20% increase in accuracy on the PASCAL VOC 2011 and 2012 challenge datasets [9, 8] compared to state of the art. They have also been used to tackle several computer vision tasks, including the segmentation of medical images [75, 55].

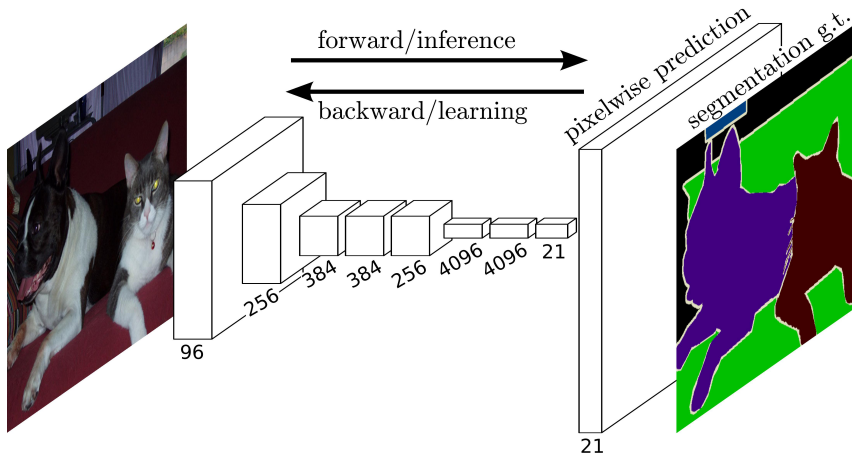


Figure 2.5: FCN architecture. Source: [54]

2.3.5 Convolutional Models with Graphical Models

Another recent innovation in Deep Learning for Image Segmentation is the use of Probabilistic Graphical Models. Conditional Random Fields are one such example of a graphical model. They are known to consider a whole scene (context) when predicting, making them particularly useful for analysing structured data such as images [46]. For this reason, researchers have studied their combination with CNNs, which ignore contextual information altogether, to create precise image segmentation frameworks.

Chen et al., 2016 [23] show that CNNs alone are not sufficient for image segmentation, but that their combination with fully connected CRFs achieves state-of-the-art results. In their work, the output of a CNN is used as input of the FC-CRF, after it was upsampled using interpolation.

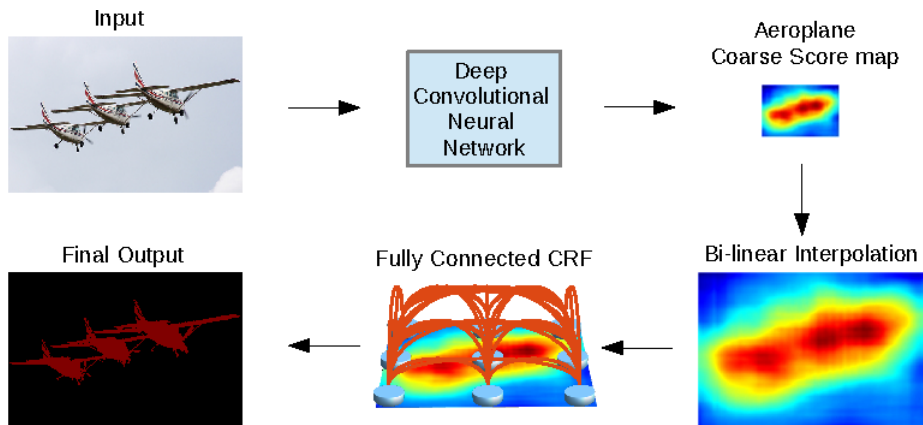


Figure 2.6: CNN + CRF architecture. Source: [23]

Schwing and Urtasun, 2015 [61] manage to combine CNN and CRF into a single trainable framework, passing the error of the CRF into the CNN. Finally, Liu et al., 2015 [53] propose an efficient implementation of the CNN+Graphical Model combination, using extra hidden layers to approximate the Mean Field algorithm for Markov Random Field learning.

2.3.6 Deep Encoder-Decoder Models

As the name suggests, Encoder-Decoder models make use of two components; an encoder which transforms the input of the model into a latent state, and a decoder, which takes this state as input and decodes it into an interpretable output. This architecture has become increasingly more popular in DL-based image segmentation, and many state-of-the-art algorithms fall into this framework [56].

Noh et al., 2015 suggest the use of a CNN, namely the VGG-16 for its uniformity, as an encoder, which uses a reversed architecture to the encoder (see Figure 2.7) [58]. Each convolutional layer in the decoder is replaced by deconvolution, and each pooling layer is replaced by unpooling.

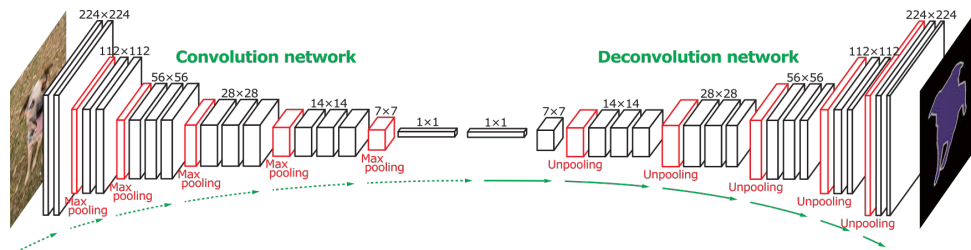


Figure 2.7: Encoder-Decoder architecture using deconvolution. Source: [58]

As described earlier, the deconvolution layer is identical to the convolutional layer, except its messages are switched. A similar relationship applies to the pooling and the unpooling layers. The output of the decoder is then the segmented image directly.

An improvement to this architecture was proposed by Badrinarayanan et al., 2015 [17]. Their SegNet architecture is similar to what was proposed by Noh et al.; however, its encoder network is fully convolutional, and it uses learnt parameters of the pooling layers with the unpooling layers.

U-net Architecture

The U-net architecture technically belongs to the Deep Encoder-Decoder Models; however, it was designed with the segmentation of high-resolution medical images in mind. For this reason, their aspects are relevant to this thesis and to potential follow-up work especially.

The U-net architecture was proposed by Ronneberger, Fisher and Brox, 2015 [60]. The motivation behind this architecture was to make use of the Fully convolutional network architecture in combination with the Encoder-Decoder structure to train a working network on only little data.

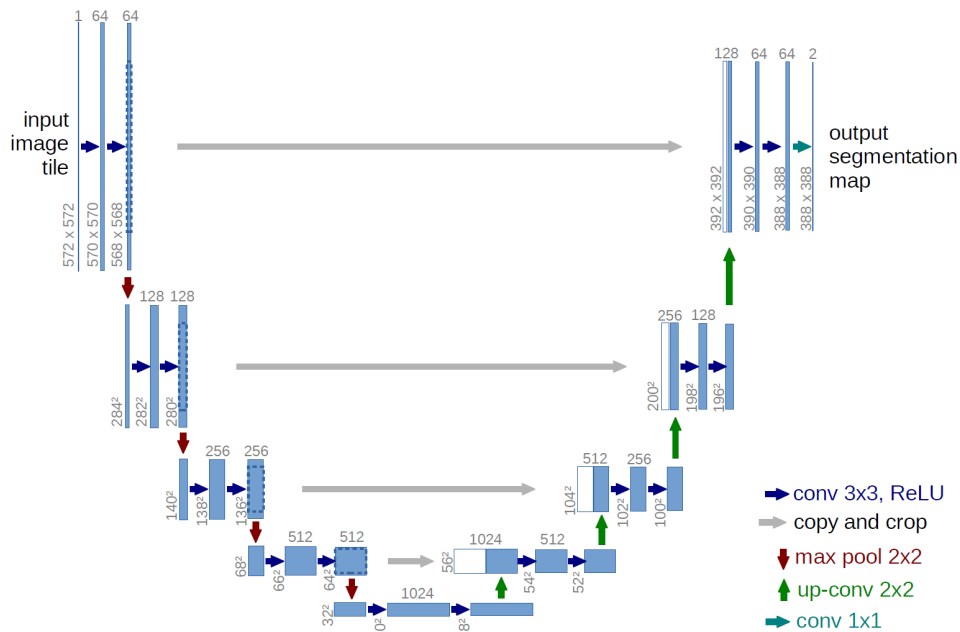


Figure 2.8: U-net architecture. Source: [60]

As is shown in the architecture schema, the U-net consists of a fully convolutional encoder and a fully convolutional decoder. Furthermore, the input to each layer of the decoder is concatenated with the output of the corresponding encoder layer, ensuring that the high-level information is not lost. The authors also make use of data augmentations during training, achieving good performance with only a small amount of training images.

The U-net has become a popular framework for image segmentation, seeing multiple applications [80, 28].



Chapter 3

Essential Algorithms and Structures

This chapter covers the principles behind the algorithms and structures used in the scope of this thesis. Namely, convolutional neural networks and algorithms from the multiple-instance learning framework are discussed. Some prior knowledge of machine learning algorithms is expected from the reader, as elementary concepts (e.g. Multi-Layer Perceptron) are not explained in full.



3.1 Convolutional Neural Networks

Convolutional neural networks make up the base of our proposed methods. To understand the used network structures, the basic layers types and other components first need to be set out. Typically, a CNN consists of alternating Convolution and pooling layers, followed by a block of Fully-connected layers. Some additional techniques can also be implemented to improve performance or the speed of convergence, such as Dropout and Batch normalisation. All these components, as well as different activation functions, are described in this section of the thesis. The basic component definitions were adopted from Khan et al., 2020 [39].

3.1.1 Convolutional Layer

Convolutional layers are the basic building block of every convolutional neural network. They consist of a set of learnable kernels (also known as filters), which come in the form of spatially small matrices (a couple of units in height and width), with a depth equal to the depth of the input (e.g. 3 for RGB images). Furthermore, each input is divided into slices, equal in size to the defined kernels. These slices are also known as receptive fields and help with the capturing of spatial features of the input. Commonly, receptive fields are chosen such that they cover the entire input, organised into a regular grid with a defined stride (distance from neighbouring receptive fields' centre locations). The kernel can then be seen as a window sliding across the input.

The outputs of the convolutional layer are computed as the dot products (sums of element-wise multiplications) between the kernels and the receptive fields. Therefore, we get one output per kernel, and each output has the size equal to the number of receptive fields (1 value per receptive field). For example, when designing a convolutional layer for inputs of size $[5 \times 5 \times 3]$, we could use two kernels of size $[3 \times 3 \times 3]$ and complete coverage of the input with nine receptive fields (organised into a 3×3 grid with stride equal to 0). The size of the output of this layer would therefore be $[3 \times 3 \times 2]$.

Formally, the convolution operation for one kernel can be expressed as:

$$f(m, n) = \sum_c \sum_i \sum_j K[i, j] F_{m,n}[i, j] \quad (3.1)$$

where $f[m, n]$ is the value of the output matrix at position $[m, n]$, c is the channel index, i, j are indices of the kernel, K is the kernel in matrix form, and $F_{m,n}$ is the receptive field in matrix form, corresponding to the position $[m, n]$ in the output matrix.

An example operation of a convolutional layer with one kernel of size $[3 \times 3]$ and input with one channel can be seen in Figure 3.1.

3.1.2 Pooling Layer

The pooling layer is a layer type designed to reduce data dimensions in CNNs. Similarly to the convolutional layer, it uses receptive fields in the form of a sliding window to split the input into regions. However, there is no learnable

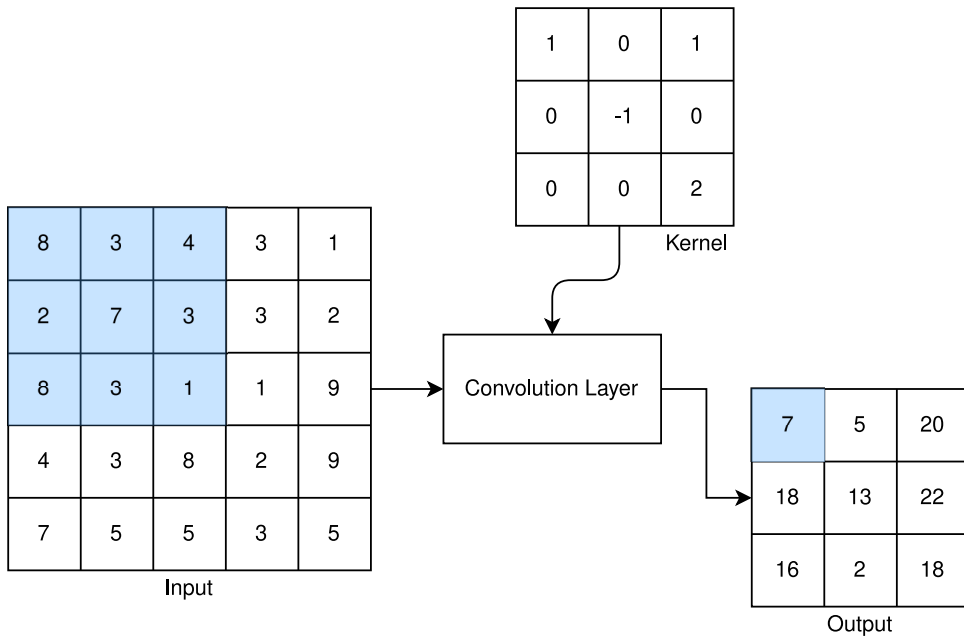


Figure 3.1: Example convolutional layer operation. One receptive field and its corresponding output value have been marked blue.

kernel as part of the pooling layer. Instead, a function is applied locally to the regions of the input covered by the receptive fields.

The most common in practice are the max-pooling layer and the average-pooling layer. As their names suggest, the max-pooling layer outputs the maximum value in each receptive field, and the average-pooling layer outputs the average. These two types of pooling layers are also used in the widely-used VGG [63] and Inception [67, 68] networks.

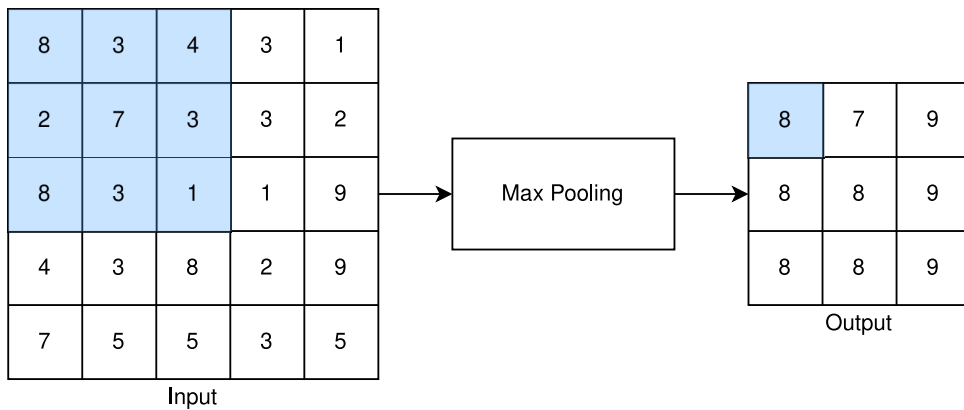


Figure 3.2: Example max-pooling layer operation. One receptive field and its corresponding output value have been marked blue.

■ 3.1.3 Fully-connected Layer

A fully connected layer connects all neurons from one layer to all neurons in the next layer. If the layer input is in multi-dimensional form, it is first flattened. Fully connected layers are typically at the end of the network and serve as classifiers that work with features extracted with the convolutional and pooling layers. The structure of the layer is the same as that of a layer in the standard Multi-Layer Perceptron algorithm.

■ 3.1.4 Activation Functions

Activation functions are an essential part of any deep neural network. Many different activation functions have been used in CNNs; however, the ReLU (Rectified Linear Unit) and its derivatives are the standards, as they deal with the vanishing gradient problem [36].

The ReLU activation function can be represented as:

$$\phi(z) = \max(0, z)$$

One of the functions derived from the ReLU function is the Leaky ReLU. It is designed to deal with the "Dying ReLU" problem. The Dying ReLU problem occurs, when a ReLU activated neuron always outputs 0 due to its input always being negative. The Leaky ReLU solves this problem by outputting small negative values for negative inputs.

The Leaky ReLU activation function can be represented as:

$$\phi(z) = \max(\alpha z, z)$$

where α is a small positive constant (e.g. 0.01).

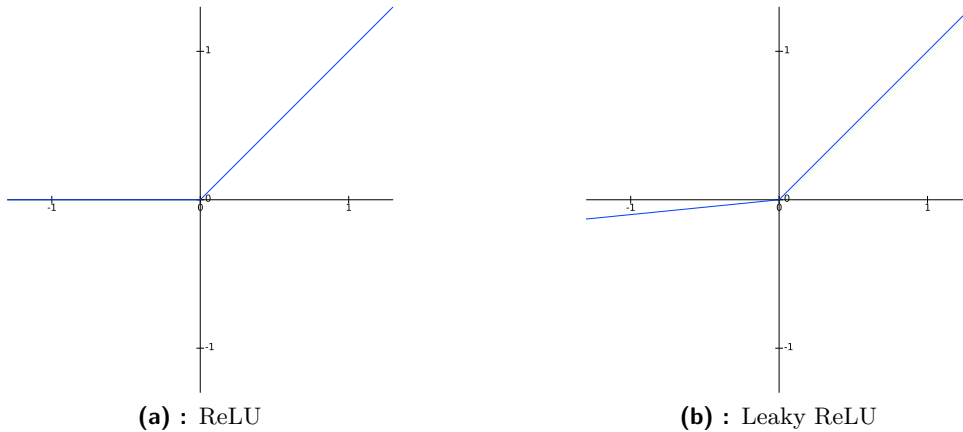


Figure 3.3: Standard CNN activation functions.

Frequently, the Softmax activation function is also used in CNNs. This activation function takes a vector of values (scores) as input and outputs values that are positive and sum up to 1, which can be interpreted as probabilities. For this reason, it is regularly used with the last layer of a neural network.

The Softmax activation function can be expressed as:

$$\phi_i(\vec{z}) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}}$$

where \vec{z} is the input of the Softmax function in the form of a vector of length n .

3.1.5 Batch-normalisation

Batch normalisation was first introduced by Ioffe and Szegedy, 2015 [38] to increase the speed of convergence of the Inception network. It normalizes the input values for each mini batch by setting their distributions to zero mean and unit variance. It can be expressed as:

$$\hat{z}_i = \frac{z_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

,where μ_B is the mean of input values for a mini batch, σ_B is the standard deviation of input values for a mini batch, ϵ is a small constant value, added for numerical stability.

3.1.6 Dropout

Dropout is a technique used to prevent overfitting in deep neural networks. It works by temporarily removing units from the network. The removed units are chosen at random at each iteration of training [65].

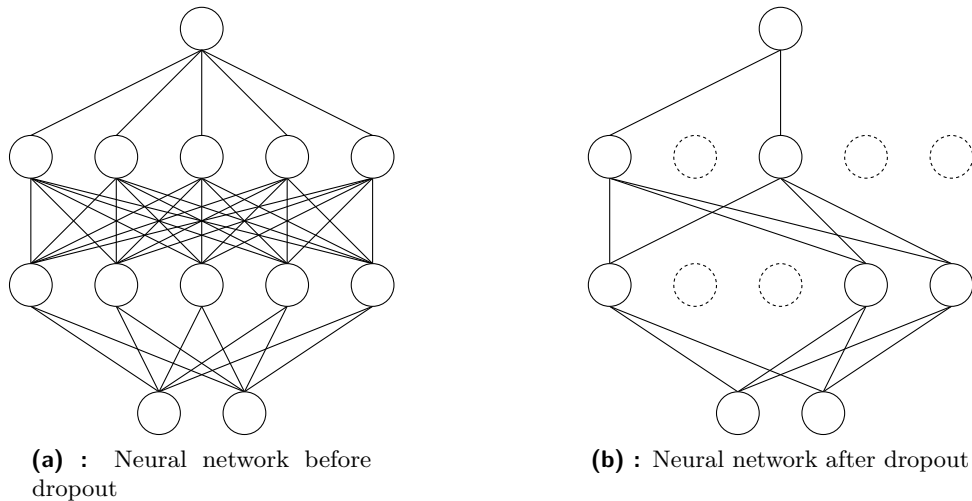


Figure 3.4: A visualisation of Dropout.

3.1.7 Inception Architecture

In our proposed method, we adopt the InceptionV3 architecture proposed by Szegedy et al., 2015 [68]. This architecture is a direct successor to the GoogLeNet, which is an incarnation of the first Inception architecture [67]. The main motivation behind the Inception architecture was to build deeper and wider networks for image classification while keeping them computationally feasible.

The authors achieve this by using "Inception blocks", which consist of convolutional layers with kernel sizes 1x1, 3x3 and 5x5 pixels, and pooling layers, all used in parallel. To reduce the number of computations necessary, the authors also add convolutional layers with a kernel size of 1x1 before each 3x3 and 5x5 convolutional layer, as well as after every pooling layer. The authors show that this reduces the number of parameters by a factor of 10 while experiencing no decrease in performance. The outputs from all parts of the Inception blocks are concatenated. The whole network is then created by connecting many inception blocks in a series. A schema of the Inception

block can be observed in Figure 3.5.

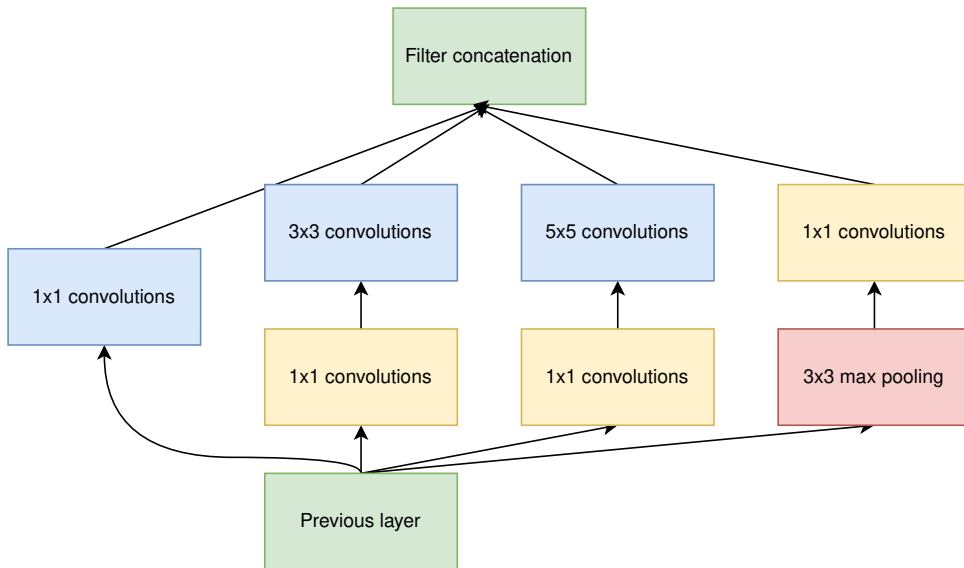


Figure 3.5: Schema of the Inception block. Adapted from [67].

3.2 Multiple Instance Learning Algorithms

Multiple instance learning (MIL) is a type of weakly supervised learning. Instead of processing training instances separately, they are grouped together to form "bags". In this case, the instance-level labels are hidden; only the label of the bag is known. Most algorithms from this framework work with the standard MIL assumption, which can be expressed as:

- All instances (feature vectors) have a hidden label $y_i \in \{-1, +1\}$
- A bag is labelled positive if at least one of its instances has a positive label.
- A bag is labelled negative if all its instances have a negative label.

Two algorithms belonging to this framework are used in the scope of this thesis: mi-SVM [15] (short for multiple instance support vector machines) and a ratio-constrained multiple instance Markov network (RMIMN) [30].

3.2.1 mi-SVM

The mi-SVM algorithm was proposed by Andrews et al., 2002 [15]. It also works with the aforementioned standard MIL assumption.

The mi-SVM algorithm works by iteratively applying the SVM algorithm to predict the hidden instance labels (i.e. labels are imputed to each instance). If all instances in a positive bag are given negative labels, the instance closest to the separating hyperplane of the SVM is labelled positive, to satisfy the MIL assumption. The pseudocode of the algorithm can be seen in Figure 3.6.

```

initialize  $y_i = Y_I$  for  $i \in I$ 
REPEAT
  compute SVM solution  $w, b$  for data set with imputed labels
  compute outputs  $f_i = \langle w, x_i \rangle + b$  for all  $x_i$  in positive bags
  set  $y_i = \text{sgn}(f_i)$  for every  $i \in I$ ,  $Y_I = 1$ 
  FOR (every positive bag  $B_I$ )
    IF ( $\sum_{i \in I} (1 + y_i) / 2 == 0$ )
      compute  $i^* = \arg \max_{i \in I} f_i$ 
      set  $y_{i^*} = 1$ 
    END
  END
  WHILE (imputed labels have changed)
  OUTPUT ( $w, b$ )

```

Figure 3.6: Pseudocode of the mi-SVM algorithm. Source: [15].

3.2.2 Ratio-constrained Multiple Instance Markov Network (RMIMN)

The use of Markov networks for multiple instance learning was proposed by Hajimirsadeghi and Mori, 2015 [30]. The schema of their proposed network can be seen in Figure 3.7. The authors define a scoring function for this network as a sum of several potentials. One potential is the instance-label potential, which is between each feature vector (instance) and its label. This potential is labelled ϕ_w^I in the schema. The second potential is the labels-clique potential, which is between all the instance labels and the bag label, denoted by ϕ_w^C . Finally, a bag-label potential is defined, which is the potential between the bag label and some feature vector \mathbf{X} , which describes the whole bag. The bag-label potential is labelled ϕ_w^B .

We adopt the Markov network architecture in our proposed method. In

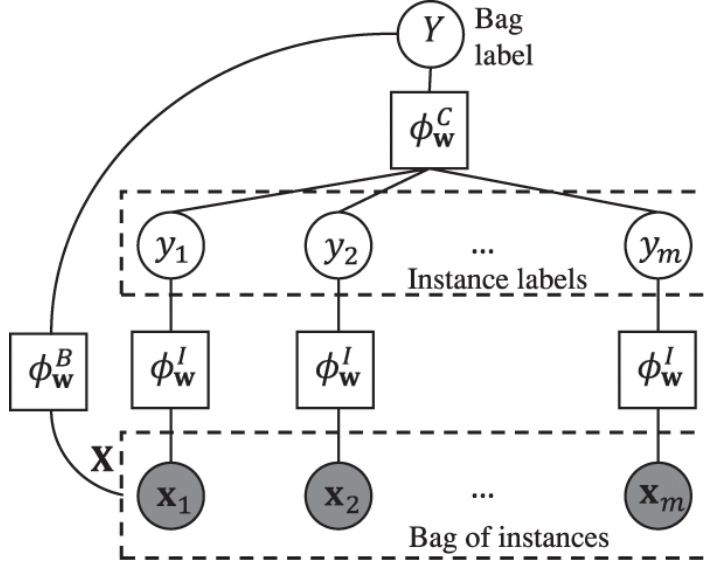


Figure 3.7: Schema of the proposed RMIMN model. Source: [30].

our case, we have no feature vector \mathbf{X} describing bags as a whole, and for this reason, we do not define a bag-label potential. The scoring function of our network can then be expressed as:

$$f_w(\mathbf{x}, \mathbf{y}, Y) = \sum_{i=1}^m \phi_w^I(\mathbf{x}_i, y_i) + \phi_w^C(\mathbf{y}, Y) \quad (3.2)$$

where m is the size of the bag.

Similarly to the authors [30], we define the potential function between one feature vector \mathbf{x}_i and its label y_i as:

$$\phi_w^I(\mathbf{x}_i, y_i) = y_i(\mathbf{w}^T \mathbf{x}_i + b) \quad (3.3)$$

,where \mathbf{w} is a vector of learnable weights and b is the learnable intercept point (bias).

Next, we define the labels-clique potential. This potential represents the multiple instance learning assumption. For this model, we step away from the standard MIL assumption, and instead, we use the generalised MIL assumption. Algorithms from the generalised MIL framework have been shown to improve the prediction accuracy when applied to similar datasets [34]. We denote the number of instances with a positive label in a bag by m^+ and the number of all instances in a bag by m . Given a constant $\rho \in (0, 1)$, we can express the generalised MIL assumption as:

- All instances (feature vectors) have a hidden label $y_i \in \{-1, +1\}$

- A bag is labelled positive iff. $\frac{m^+}{m} > \rho$.

To represent this assumption in the Markov network, two cardinality potential functions are used, one for bags with a positive label (denoted by $C_{\mathbf{w}}^+$), the other for bags with a negative label (denoted by $C_{\mathbf{w}}^-$). These functions are defined as:

$$\begin{aligned} C_{\mathbf{w}}^+(m^+, m) &= -\infty & 0 \leq \frac{m^+}{m} < \rho \\ C_{\mathbf{w}}^+(m^+, m) &= 0 & \rho \leq \frac{m^+}{m} < 1 \\ C_{\mathbf{w}}^-(m^+, m) &= 0 & 0 \leq \frac{m^+}{m} < \rho \\ C_{\mathbf{w}}^-(m^+, m) &= -\infty & \rho \leq \frac{m^+}{m} < 1 \end{aligned}$$

The optimal value of ρ is found experimentally. This value is likely to be different for each attribute.

Finally, the labels-clique potential is defined as:

$$\phi_{\mathbf{w}}^C(\mathbf{y}, Y) = C_{\mathbf{w}}^+(m^+, m)(Y=1) + C_{\mathbf{w}}^-(m^+, m)(Y=-1) \quad (3.4)$$

Using a labels-clique potential designed in this way, our scoring function outputs $-\infty$ for every labelling that does not satisfy the generalised MIL assumption.

■ Inference

The inference problem is to find the optimal values of \mathbf{y}_i given the bag-level label Y . These values can be found by solving:

$$\mathbf{y}^* = \max_{\mathbf{y}} \sum_{i=1}^m \phi_{\mathbf{w}}^I(\mathbf{x}_i, y_i) + \phi_{\mathbf{w}}^C(\mathbf{y}, Y) \quad (3.5)$$

An efficient algorithm is described in [30], which solves the problem in $\mathcal{O}(m \log m)$ time. We adopt this algorithm to solve the inference problem for our model. We implement it as follows:

- Sort all instances according to $\phi_{\mathbf{w}}^I(\mathbf{x}_i, +1)$
- For $k = 0, \dots, m$, compute $s_k = \sum_{i=1}^k \phi_{\mathbf{w}}^I(\mathbf{x}_i, +1) + C_{\mathbf{w}}^+(k, m)(Y=1) + C_{\mathbf{w}}^-(k, m)(Y=-1)$

- Choose k^* for which s_k is maximal
- Label top k^* instances as positive, label rest of the instances as negative

■ Training

The RMIMN is trained by alternating between inference and learning. During inference, the instance labels are predicted. Next, the instance-label potential function weights and bias are learned by running a standard SVM algorithm on all the instances together, using the predicted instance labels as the ground truth. A linear kernel is used for the SVM, the weights and bias of which are used directly as the weights and bias of the instance-label potential function after each learning step. The algorithm is run until there is no change in the inferred instance labels.



Chapter 4

Methodology

In this chapter, the methodology of the thesis is discussed.

In the first part of this chapter, the proposed method for segmenting WSI images into normal tissue, tumour tissue and background is described. Firstly, the dataset is described in more detail. Secondly, the proposed method of segmentation, using a convolutional neural network, is outlined. Next, a method of distinguishing tissue and background is talked about. Then, the means of patch extraction is reported. Afterwards, the design and aspects of the proposed patch-classifier architecture are described. Finally, the training of the classifier is talked about.

In the second part of the chapter, the proposed methods of classification according to three attributes is described. Both methods make use of the patch method. Firstly, we propose a method from the MIL framework, treating each image as a bag of extracted patches. Secondly, we propose a method of classifying each patch independently.



4.1 PETACC3 Dataset

As has been reported, The PETACC3 trial was a phase III trial (a trial where the new drug is compared to the standard-of-care drug) with random assignment of treatment in multiple medical centres to test the effectivity

of the addition of irinotecan to other forms of medication for patients with stage 3 colorectal cancer [3]. The working dataset for this thesis consists of microscopy images scanned during this trial. It spans 1140 images belonging to 28 anonymous patients (30 to 54 images per patient), each labelled on the image-level by an expert pathologist. The labelling was done across 3 attributes: mucinous, serrated and Crohn-like.

The labels are defined as follows:

- Mucinous class: 1=no; 2=minimal; 3=moderate ; 4=yes (>50%)
- Serrated class: 1=no; 2= minimal; 3=moderate; 4=abundant
- Crohn-like class: 1=yes; 2=no; 3= not able to determine

Furthermore, the expert marked 150 images as "pure-case" with respect to the mucinous attribute. These images are well-defined without many unwanted artefacts (e.g. tissue tears) and with an abundance of class-specific indicators (i.e. mucus) or their complete absence.

Additionally, pixel-level annotations are provided for most images. These annotations were created to distinguish areas with tumorous cells. For this reason, they are used as ground truth for the tumorous/normal tissue segmentation.

From the example image in Figure 4.1, it is clear, that images in the dataset contain many marker lines. Furthermore, a large part of the image is the background, which should not be considered when training a classifier. The means of removing the artefacts and segmenting the image is described in the next sections of this chapter.

■ 4.2 WSI Segmentation Methodology

When designing the method for WSI segmentation for this part of the thesis, a few assumptions were made.

The first assumption is that when segmenting images into background and tissue, the high-resolution information (e.g. cell shape) is not necessary. The



Figure 4.1: Example image from the PETACC3 dataset. Marked pure-case with labels: mucinous=1, serrated=2, Crohn-like=2.

feasibility of this assumption has been shown by Hering and Kybic, 2020 [34]. This allows the segmentation of this type to be sufficiently precise when using down-sampled images.

The second assumption is that the segmentation of tissue into normal and tumour tissue only requires small-scale information. It is further assumed, that the necessary low-level signal is fully recognisable in any sufficiently large area. In other words, the image can be segmented by classifying sufficiently large patches extracted from it. Using this patch method, we lose the large-scale information, which is assumed to be unnecessary. This has been shown to be a feasible method for tumor/normal tissue segmentation, when pixel-level annotations are available, which is also the case in this thesis [16, 18].

Using these assumptions, a method of segmentation into normal tissue, tumour tissue and background is proposed as follows: Firstly, segmentation masks are created, by segmenting down-sampled WSIs into background and tissue. Secondly, these masks are applied to extract sufficiently large patches, forming a dataset by labelling each patch according to the provided pixel-level annotations. Thirdly a patch classifier is designed, implemented and trained on these labelled patches. Finally, WSI segmentation is achieved by labelling each pixel according to the corresponding patch label.

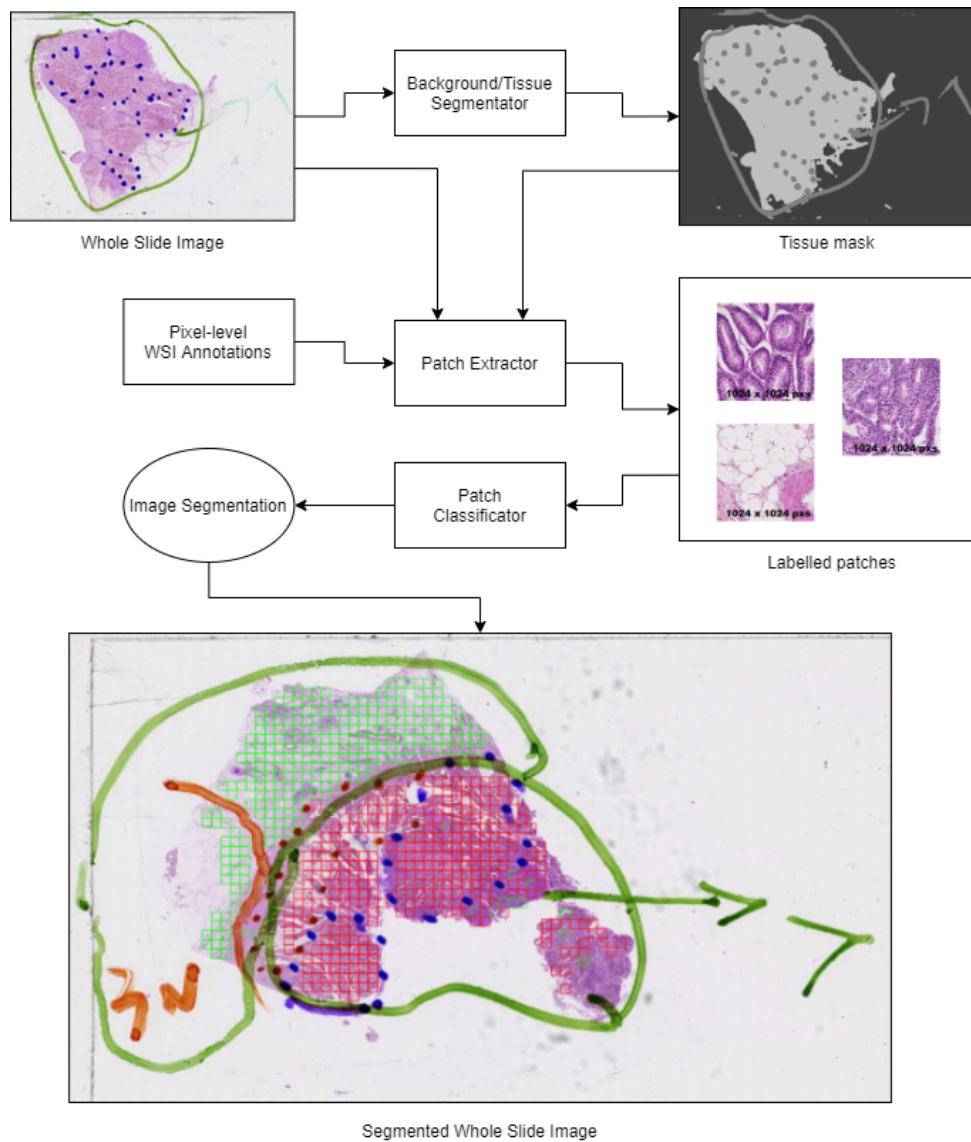


Figure 4.2: Schema of the entire proposed WSI segmentation method.

4.3 Segmentation into Tissue, Background and Pen Marks

We use a random forest classifier to segment down-sampled images from the dataset. We use the Ilastik image segmentation software [20] to create a tissue/background/pen-mark pixel mask for each image of the dataset. The weights of the classifier are provided by the thesis supervisor, using the same values as in Hering and Kybic, 2020 [34], so no additional training was

necessary. An example segmentation can be found in Figure 4.3.

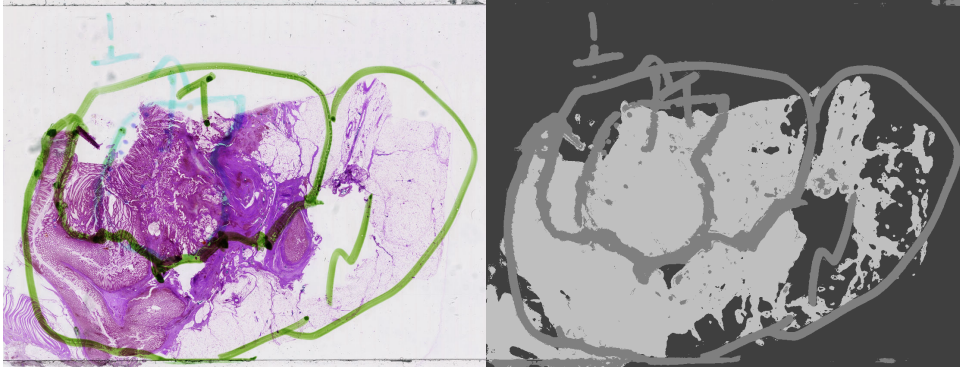


Figure 4.3: Example Background/Tissue segmentation. Original down-sampled image on the left, segmentation mask on the right. Tissue in light grey, marker lines in darker grey, background in dark grey.

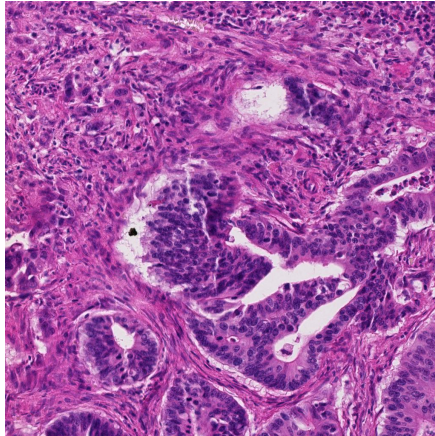
4.4 Patch Extraction

By using the expert annotations and the generated tissue masks, patches were then extracted from the WSI. The patch size of 1024x1024 pixels was chosen from early validation. Only patches containing at least 80% of tissue were chosen for further use, the rest was omitted. The patch extraction was done using software provided by the group of Biomedical imaging algorithms at the Czech Technical University in Prague.

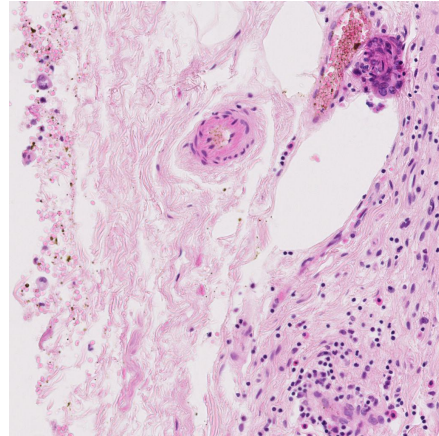
During patch extraction, data was split into 75% training data, 10% validation data and 15% testing data. As patches were created in a per-image manner, it was easy to ensure no patient was used in two data splits simultaneously, therefore making it easier to avoid potential overfitting. Each output patch is labelled either TU/NO depending on the annotation of the area it originated from. The TU label means there are tumorous cells in the annotated area, the NO label means there are none.

4.5 Model Architecture and Training

We use a convolutional neural network as our patch classifier, specifically the InceptionV3 architecture which was described in section 3.1.7. The training of the Inception model was done using the PyTorch library, which is a deep



(a) : Patch labelled "TU"



(b) : Patch labelled "NO"

Figure 4.4: Example extracted patches from the PETACC 3 dataset. Patches were labelled TU and NO respectively. Both patches have dimensions of 1024x1024 pxs.

learning library, designed mainly for the Python programming language [59]. It provides efficient implementation of deep learning techniques, including various optimisation algorithms. It relies on the use of tensors, optimised for use with graphics processing units. Furthermore, the implementations of many neural network architectures are included in this library, including the InceptionV3 architecture. We use the InceptionV3 model pretrained on the ImageNet [25] database, to make use of transfer learning, which has been shown to significantly increase speed of convergence in some cases [73].

4.5.1 Input Transformations

During training, the input data was downscaled to the size of 299x299 pixels, which is the input size of the InceptionV3 network. Next, the input images are randomly flipped, and normalised with $\mu = (0.485, 0.456, 0.406)$ and $\sigma = (0.229, 0.224, 0.225)$ which are the mean and standard deviation values of the ImageNet dataset, on which the Inception network was pre-trained. Finally, the input values were transformed from the RGB (Red, Green, Blue) spectrum to HSV (Hue, Saturation, Value), as suggested in Halcek et al., 2019 [31]. The authors show, that such a transformation can lead to a performance increase in WSI segmentation using CNNs.

4.5.2 Network Parameters and Training

For the purpose of binary classification of patches into tumorous and regular tissue, the output layer of the standard InceptionV3 model was replaced with an output layer of only one neuron.

Binary softmax cross entropy loss was chosen as the loss function, which is the standard loss function for binary classification [76]. It combines the softmax activation function described in section 3.5.4 with the binary cross entropy loss function. The binary cross entropy loss function can be expressed as:

$$CE = -\frac{1}{m} \sum_{i=1}^m y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log (1 - \hat{y}_i)$$

where y is the target label, \hat{y} is the label predicted by the classifier and m is the number of instances.

The loss function was optimised using the stochastic gradient descent method. The initial learning rate of the network was set to 0.1, reduced by a factor of 10 on plateau (i.e. when the validation error no longer decreases). The minimal learning rate was set to 10^{-5} .

The network was trained on a Nvidia RTX 2070 GPU. The training was done over a maximum of 30 epochs, with early stopping if no further improvement to the validation error was detected and the minimal learning rate was reached.

4.6 WSI Classification Methodology

In this section of the thesis, the method of classifying whole WSIs from the PETACC3 dataset according to tumour type is discussed. The WSIs are to be classified according to three attributes: mucinous, serrated and Crohn-like. These attributes are not mutually exclusive; an image can therefore be classified as any combination of the three.

Furthermore, numerical labels are provided for each image, rather than binary labels (see Section 4.1). Consequently, we threshold these expert labels to create a positive and negative class for each attribute. The following thresholding was decided:

Attribute	Values for Negative Class	Values for Positive Class
Mucinous	{1, 2}	{3, 4}
Serrated	{1, 2}	{3, 4}
Crohn-like	{2}	{1}

Accordingly, the task can be thought of as three independent cases of binary classification.

The images have only image-level labels with respect to these attributes, therefore the assumption, that the signal is fully recognisable on small patches extracted from the image, that was made when segmenting images according to tumorous and normal tissue might not be feasible. For this reason, we treat the classification problem as a multiple instance learning task.

4.7 MIL Approach

The motivation behind the MIL approach is to treat the given WSIs as bags of unique tissue patches. We apply the same tissue/background/pen mark segmentation and patch extraction as when segmenting the images according to tumour type (see Section 4.3 and Section 4.4). However, the size of the generated patches is too large to process entire bags at once. For this reason, we need to design a patch descriptor to reduce the dimensionality of the patches. The feature vectors generated from the patches using this descriptor are then classified as bags using MIL algorithms.

4.7.1 Patch Description

To reduce the dimensionality of the generated patches, we propose to train a convolutional neural network on them to be used as a patch descriptor. For this purpose, we use the InceptionV3 architecture, with the output layer of the network replaced by two fully connected layers with 32 neurons and 1 neuron respectively. The network is trained to classify patches into tumorous and normal, using the training process from Section 4.5. The last layer serves to learn the weights of the network and is removed after training. We are then left with a CNN with 32 outputs, which we assume to sufficiently describe the information from each patch. Finally, 32 features are extracted from all the patches, that were labelled as tumorous, using this network. We omit

the patches extracted from normal tissue, as they would add noise to the classification task (normal tissue can not belong to any of the attributes).

■ 4.7.2 Bag Classification

The features that were extracted (by the patch description step) from patches belonging to the same image are grouped together to form "bags". The bags are then labelled either positive or negative for each attribute, according to their expert labels. We then use algorithms from the multiple instance learning framework to classify these bags and consequently classify the whole WSIs (the bag label is equal to the image label).

Two algorithms are adopted for this purpose: mi-SVM (described in Section 3.2.1) and ratio-constrained multiple instance Markov network. (described in Section 3.2.2).

■ mi-SVM

We implement the mi-SVM algorithm, according to Andrews et al., 2002 [15]. We first evaluate our implementation on known MIL datasets, then apply it to the bags of described patches. We adapt the standard MIL assumption for our case. By an instance, we mean the feature vector extracted from a patch by the patch descriptor. We formulate the MIL assumption as:

- All instances (and their respective patches) have a hidden label $y_i \in \{-1, +1\}$
- A bag (and its respective image) is labelled positive if at least one of its instances has a positive label.
- A bag is labelled negative if all its instances have a negative label.

The pseudocode of the implemented mi-SVM algorithm, as well as other details can be seen in Section 3.2.1.

■ Ratio-constrained Multiple Instance Markov Network

We implement the RMIMN algorithm, proposed by Hajimirsadeghi and Mori, 2015 [30]. We implement a simplified means of optimisation, using a linear kernel SVM to learn the network parameters. We adapt the generalised MIL assumption when implementing the algorithm.

We denote the number of instances with a positive label in a bag by m^+ and the number of all instances in a bag by m . Given a constant $\rho \in (0, 1)$, we can express the generalised MIL assumption for our case as:

- All instances (and their respective patches) have a hidden label $y_i \in \{-1, +1\}$
- A bag (and its respective image) is labelled positive iff. $\frac{m^+}{m} > \rho$.

We learn the optimal value of ρ experimentally, independently for each attribute.

■ 4.7.3 Schema of the MIL Approach

The entire proposed MIL approach is illustrated in Figure 4.5

■ 4.8 Direct Approach

For the direct approach, the same model, which was described in Section 4.5, is used. Unlike the MIL approach, where we acquire image-level labels, we classify the images at the patch-level. For this purpose, we use the InceptionV3 network, with the last layer replaced by a fully-connected layer with three neurons, one for each attribute. The network was trained using the stochastic gradient descent algorithm, with the same hyperparameters. The same transformations were applied to the input.

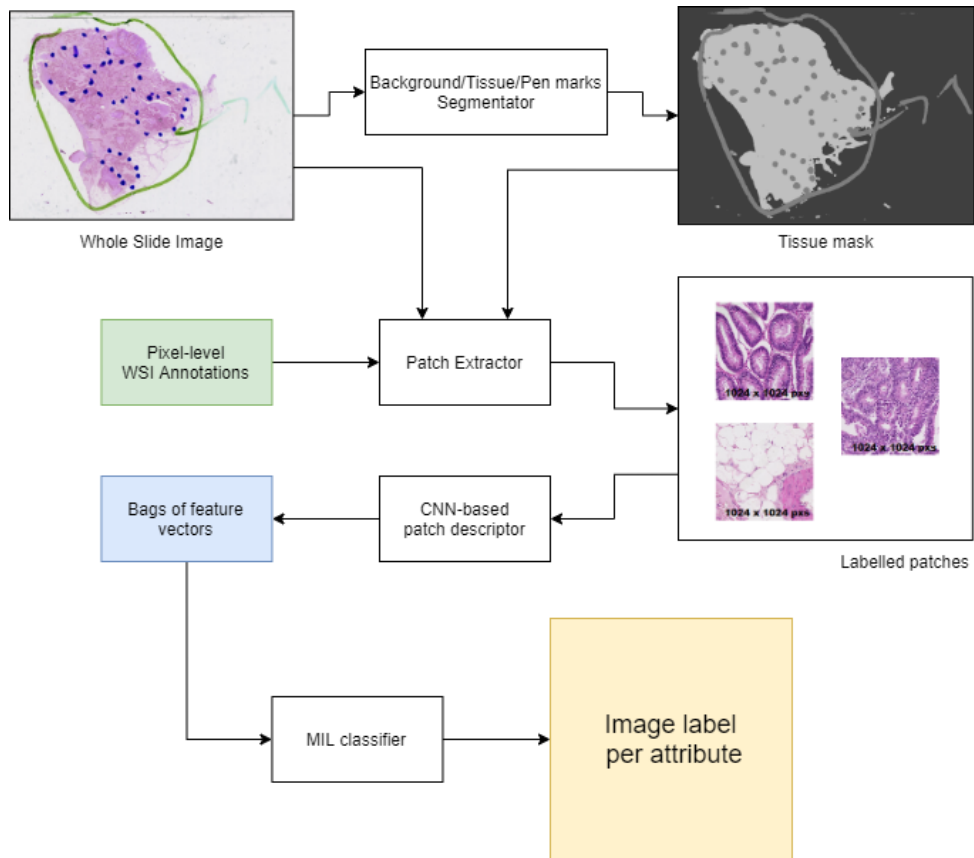



Figure 4.5: Schema of the proposed MIL approach.



Chapter 5

Experiments

The performed experiments are discussed in this chapter. For each experiment, we describe how the experiment was done, the purpose of the experiment, as well as the results and their discussion.



5.1 Experiment 1 - WSI Segmentation into Tumorous and Normal Tissue, CNN Trained on Pure-case Images

In this experiment, we test the implemented CNN architecture on the PETACC3 dataset. The CNN is trained and evaluated only on the images the expert marked as "pure case" (see Section 4.1). 15% of the data is used for testing. The test set consists of histopathological images, taken from patients that are not a part of the training and validation sets. Each classifier is trained for a maximum of 30 epochs, stopping early if there is no further improvement to the validation error.

The purpose of this experiment is to evaluate the correctness of the CNN implementation, as well as its performance on the PETACC3 dataset, given a smaller amount of high-quality data (about 1/6th of the whole dataset is used). The performance of the segmentation is measured with patch-level classification accuracy (e.g. 90% segmentation accuracy implies 90% of the generated patches were classified correctly).

5.1.1 Results

The trained classifier achieved the following segmentation performance:

Accuracy	Precision	Recall
91.33%	94.99%	89.80%

We observed an accuracy of 91.33% which we consider sufficient. We also see a much higher precision (94.99%) than recall (89.80%). This could be caused by the expert annotating larger areas as tumorous than is necessary. This is common in practice, as it is often important for all tumorous tissue to be included in an annotated area. Therefore, it is possible for some extracted patches, with a positive label to, in fact, have no tumorous tissue. To better evaluate the results of this experiment, a ground truth label for each extracted patch would be necessary.

5.1.2 Example Segmentations

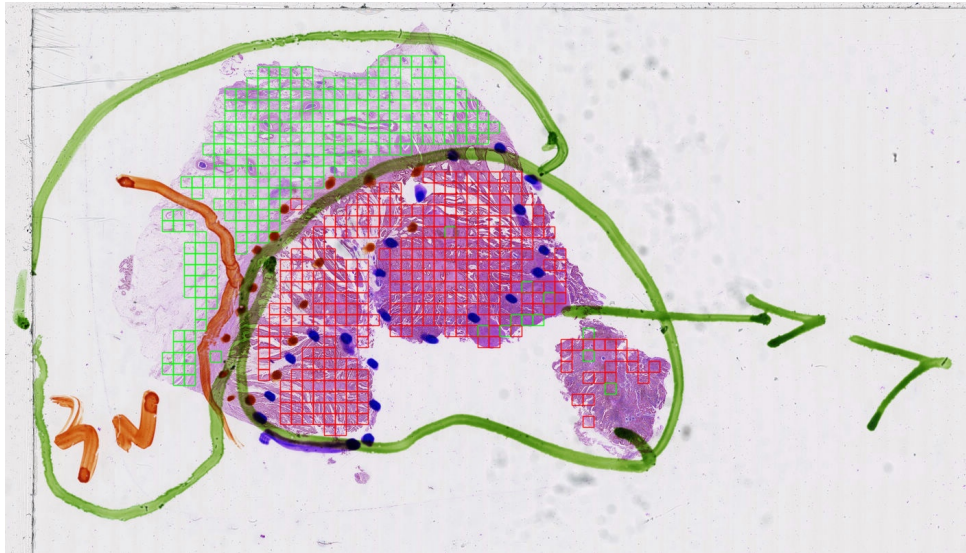


Figure 5.1: Example Normal/Tumorous tissue segmentation. Tissue patches classified as normal tissue are marked with a green border. Patches classified as tumorous are marked with a red border. The image is from the pure-case set.

We can see an example of WSI segmentation in Figure 5.1. Area annotated "N" means the area has a normal expert label. Area annotated as "T" means the area has a "tumorous" expert label. We can see 9 misclassified tumorous

tissue patches and 1 misclassified normal tissue patch. This supports the low measured recall of the classifier, compared to its accuracy.

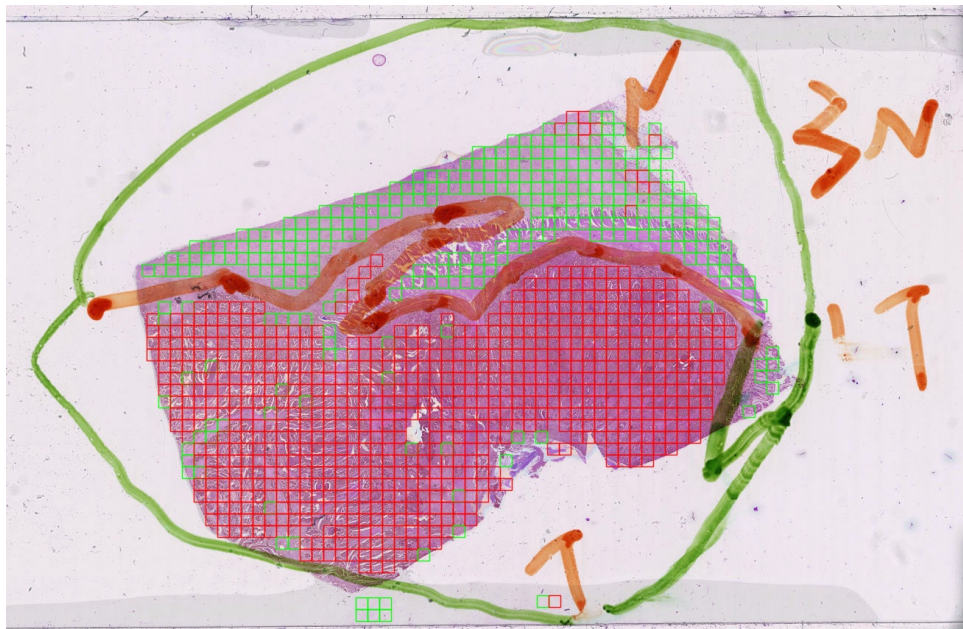


Figure 5.2: Example Normal/Tumorous tissue segmentation. Tissue patches classified as normal tissue are marked with a green border. Patches classified as tumorous are marked with a red border. The image is from the pure-case set.

Another example WSI segmentation can be observed in Figure 5.2. We see a lower accuracy than in the image in Figure 5.1. However, it is still sufficiently segmented.

5.2 Experiment 2 - WSI Segmentation into Tumorous and Normal Tissue, CNN Trained on All Images

In this experiment, we train the Inception network on patches extracted from all images from the PETACC3 dataset. The purpose of this experiment is to decide whether using all images from the dataset increases the performance of the classifier on the test set.

■ 5.2.1 Results

The classifier achieved an accuracy of 93.43%. This suggests that the use of the entire dataset increases the performance of the trained classifier. However, it is important to note, that, due to error, some images in the training and test set share the patient they originate from. This could have increased the performance of the classifier on the test set. Due to time constraints, it was not possible to remake this experiment in time, as training takes a lot of time on such a large dataset.

■ 5.3 Experiment 3 - WSI Classification According to Tumour Type - Direct Method

For this experiment, we train a CNN to classify extracted patches from the WSI according to the mucinous, Crohn-like and serrated attributes and evaluate it on the test set. We assume each extracted tumorous tissue patch from the image has the same labels as the image. We use "pure-case" images to train the classifier. The purpose of the experiment is to evaluate the feasibility of the aforementioned assumption. Since the train and test sets consist of "pure-case" images with respect to the mucinous class, we expect better performance with respect to this attribute.

■ 5.3.1 Results

The classifier achieved the following per-patch accuracy results:

Mucinous	Serrated	Crohn-like
81.30%	63.58%	52.43%

It is difficult to make conclusions from the results of this experiment, as we apply the same assumption that the image label is fully recognisable in all its extracted patches also during evaluation. However, the expectation that the classification accuracy is best for the mucinous attribute holds true. We conclude the classification accuracy is not sufficiently high for the serrated

and Crohn-like attributes, and the assumption is therefore not feasible for these attributes. For the mucinous class, the results are not conclusive.

5.4 Experiment 4 - mi-SVM trained on MUSK and MUSK2 datasets

In this experiment, we evaluate the correctness of the implemented mi-SVM algorithm by training it on the MUSK and MUSK2 datasets. These are two datasets from the UCI ML Repository [27] which are often used to evaluate algorithms from the MIL framework. We use 10-fold cross-validation when obtaining the accuracy results, averaging over the results. We compare the results with the original paper proposing the mi-SVM algorithm [15].

5.4.1 Results

We acquired the following accuracy results on the two datasets:

	MUSK1	MUSK2
Our Results	83.3%	79.8%
Cited Paper	87.4%	83.6%

We achieved similar results to the paper proposing the mi-SVM algorithm. The performance of our implementation is, nevertheless, worse than the cited paper. This could be caused by generating different validation folds or by suboptimal hyperparameters of the algorithm. We conclude our mi-SVM implementation is correct.

5.5 Experiment 5 - WSI Classification According to Tumor Type - mi-SVM

For the fifth experiment, we train the mi-SVM algorithm on extracted bags of features from the "pure-case" images of the PETACC3 dataset (see Section 4.7

for the used method) to classify them according to the mucinous, serrated and Crohn-like attributes. The accuracy results are measured at the image level. The purpose of this experiment is to evaluate whether the standard MIL assumption is feasible for bags of features defined this way. We use the Linear SVM kernel with $C=100$. We evaluate using 10-fold cross-validation.

■ 5.5.1 Results

We obtained the following image-level accuracy results for the given attributes:

Mucinous	Serrated	Crohn-like
70.45%	62.33%	54.54%

We can see that the performance of the mi-SVM classification method is not exceptional. This might be caused by the standard MIL assumption being too strict, or by a badly designed method of feature extraction; the features extracted from training the CNN to classify tissue into tumorous and normal might not sufficiently describe the patches for the purpose of classification according to tumour type. We hope to evaluate this in the last experiment.

■ 5.6 Experiment 6 - WSI Classification According to Tumor Type - RMIMN

For the last experiment, we train the implemented Ratio-constrained multiple instance Markov network on the extracted bags of feature vectors. The experiment is done to decide, whether by generalising the MIL assumption, we achieve a higher classification accuracy. We use the same training and testing sets as in Experiment 5. We again evaluate using 10-fold cross-validation.

■ 5.6.1 Results

We obtained the following image-level accuracy results for the given attributes:

Mucinous	Serrated	Crohn-like
77.81%	64.53%	56.27%

The optimal ratio of positive instances to all instances ρ for the generalised assumption was found to be 0.3 for all three attributes.

We can see a substantial increase in accuracy for the mucinous attribute. While the accuracy also increases for the other two attributes, the results still remain poor. We conclude that generalising the MIL assumption increases the accuracy of classification; however we were not able to train a satisfactory classifier for the serrated and Crohn-like attributes.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, we provided a thorough report of the state-of-the-art methods of image segmentation and classification, with a focus on the processing of whole slide images. Then we introduced and described several algorithms, which we then implemented and used to segment histopathological images into tumorous and normal tissue and to classify these images according to three attributes: mucinous, serrated and Crohn-like.

We achieved satisfactory accuracy results on the task of segmenting the WSIs into tumorous and normal tissue, which point at a correct choice of method for this task as well as a correct implementation of said methods.

For the task of classifying the pathological images according to three attributes, we achieved poor results for two of the three attributes. Several issues might be causing these poor results:

- Features extracted from a network trained to classify patches into normal and tumorous tissue might not be suited to describe the patches with respect to the attributes
- High-resolution information which is lost by splitting the images into patches might be necessary to classify the images according to the two

attributes

- The conversion of the classification problem into binary classification by thresholding the expert values. Incorrect thresholds might have been chosen when defining the task.
- The expert labels with respect to the attributes are vaguely described (1=no, 2=minimal, 3=moderate; 4=yes). This could cause some images to be given wrong labels when thresholding these expert values.

6.2 Future Work

Several experiments had to be omitted due to time constraints. Due to the dimensionality of the data, experiments take up to a couple of days to finish. A CNN should be retrained on all the available WSIs from the dataset, to reevaluate the results of Experiment 2. Different CNN architectures should also be tried out for segmenting the WSIs into tumorous and normal tissue. Furthermore, more means of patch description should be evaluated and used with the proposed MIL methods (e.g. a longer feature vector).

It would also be interesting to formulate the classification task (w.r.t. the three attributes) as a regression task, predicting the actual numerical expert values instead. This could show how precisely tuned the expert values are.

Finally, a method that processes the images as a whole (instead of the patches) should be implemented. This could increase the accuracy of classification on the serrated and Crohn-like (assuming the high-resolution lost by using the patch method is necessary).



Appendix A

Attachments

The attached files are described here. We distinguish three types of files.

- Files marked "I" were implemented from scratch (using imported libraries).
- Files marked "A" were provided by the thesis supervisor, however, changes were made to these files for the use in this thesis.
- Files marked "U" were provided by the thesis supervisor and were used completely unchanged.

Some program files are dependent on other files that fall into the CMP group framework and are not included as attachments with the thesis. Please

contact the thesis supervisor to gain access to these files if necessary.

Filename	Description	Label
train_model.py	Implements training of the CNN network.	I
create_dataset.py	Generates a dataset of patches from WSIs	I
create_bag_dataset.py	Generates a dataset of bags of patches	I
create_patch_descriptions.py	Generates feature vectors from patches	I
evaluate_model.py	Generates performance measures of a model	A
patch_extraction.py	Extracts patches from a WSI	U
petacc_patches_dataset.py	Dataset representation for use with CNN	I
classification.py	Classify a folder of WSIs	A
mil.py	Implemented MIL methods	I



Appendix B

Bibliography

- [1] Dataset-ICIAr 2018-Grand Challenge. <https://iciar2018-challenge.grand-challenge.org/Dataset/>. Accessed: 2020-12-12.
- [2] FDA allows marketing of first whole slide imaging system for digital pathology. <https://www.fda.gov/news-events/press-announcements/fda-allows-marketing-first-whole-slide-imaging-system-digital-pathology>. Accessed: 2020-12-09.
- [3] Grand Challenge. <https://grand-challenge.org/>. Accessed: 2020-12-09.
- [4] ImageNet Large Scale Visual Recognition Competition 2012 (ILSVRC2012). <http://www.image-net.org/challenges/LSVRC/2012/results>. Accessed: 2020-19-12.
- [5] ImageNet Large Scale Visual Recognition Competition 2014 (ILSVRC2014). <http://www.image-net.org/challenges/LSVRC/2014/results>. Accessed: 2020-19-12.
- [6] ImageNet Large Scale Visual Recognition Competition (ILSVRC). <http://www.image-net.org/challenges/LSVRC/>. Accessed: 2020-19-12.
- [7] ImageNet Large Scale Visual Recognition Competition (ILSVRC2015). <http://www.image-net.org/challenges/LSVRC/2015/results>. Accessed: 2020-19-12.
- [8] The PASCAL Visual Object Classes Challenge 2011 (VOC2011). <http://host.robots.ox.ac.uk/pascal/VOC/voc2011/index.html>. Accessed: 2020-19-12.

- [9] The PASCAL Visual Object Classes Challenge 2012 (VOC2012). <http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html>. Accessed: 2020-19-12.
- [10] Summary of the HIPAA Privacy Rule. <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>. Accessed: 2020-12-12.
- [11] Whole Slide Imaging | MBF Bioscience. <https://www.mbfbioscience.com/whole-slide-imaging>. Accessed: 2020-12-09.
- [12] WSI Quality Control | Bio-MIBLab. miblab.bme.gatech.edu/research/imaging/wsi-quality-control/. Accessed: 2020-13-12.
- [13] A. Aksac, D. Demetrick, T. Ozyer, and R. Alhajj. Brecahad: a dataset for breast cancer histopathological annotation and diagnosis. *BMC Research Notes*, 12, 12 2019.
- [14] S. Al-Janabi, A. Huisman, A. Vink, R. Leguit, G. Offerhaus, F. Kate, and P. Diest. Whole slide images for primary diagnostics of gastrointestinal tract pathology: A feasibility study. *Human pathology*, 43:702–7, 09 2011.
- [15] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15, pages 577–584. MIT Press, 2003.
- [16] G. Aresta, T. Araújo, S. Kwok, S. S. Chennamsetty, M. S. K. P., A. Varghese, B. Marami, M. Prastawa, M. Chan, M. J. Donovan, G. Fernandez, J. Zeineh, M. Kohl, C. Walz, F. Ludwig, S. Braunewell, M. Baust, Q. D. Vu, M. N. N. To, E. Kim, J. T. Kwak, S. Galal, V. Sanchez-Freire, N. Brancati, M. Frucci, D. Riccio, Y. Wang, L. Sun, K. Ma, J. Fang, I. Koné, L. Boulmane, A. Campilho, C. Eloy, A. Polónia, and P. Aguiar. BACH: grand challenge on breast cancer histology images. *CoRR*, abs/1808.04277, 2018.
- [17] V. Badrinarayanan, A. Handa, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling, 2015.
- [18] P. Bandi, O. Geessink, Q. Manson, M. van Dijk, M. Balkenhol, M. Hermsen, B. Ehteshami Bejnordi, B. Lee, K. Paeng, A. Zhong, Q. Li, F. Ghazvinian Zanjani, S. Zinger, K. Fukuta, D. Komura, V. Ovtcharov, S. Cheng, S. Zeng, J. Thagaard, and G. Litjens. From detection of individual metastases to classification of lymph node status at the patient level: The camelyon17 challenge. *IEEE Transactions on Medical Imaging*, PP:1–1, 08 2018.

- [19] B. E. Bejnordi, M. Veta, P. J. van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, J. A. W. M. van der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol, O. Geessink, N. Stathonikos, M. V. van Dijk, P. Bult, F. Beca, A. Beck, D. yong Wang, A. Khosla, R. Gargeya, H. Irshad, A. Zhong, Q. Dou, Q. Li, H. Chen, H. Lin, P. Heng, C. Hass, E. Bruni, Q. J. J. Wong, U. Halici, M. Ü. Öner, R. Cetin-Atalay, M. Berseth, V. Khvatkov, A. Vylegzhanin, O. Z. Kraus, M. Shaban, N. Rajpoot, R. Awan, K. Sirinukunwattana, T. Qaiser, Y.-W. Tsang, D. Tellez, J. Annuschein, P. Hufnagl, M. Valkonen, K. Kartasalo, L. Latonen, P. Ruusuvaori, K. Liimatainen, S. Albarqouni, B. Mungal, A. George, S. Demirci, N. Navab, S. Watanabe, S. Seno, Y. Takenaka, H. Matsuda, H. A. Phoulady, V. Kovalev, A. Kalinovsky, V. Liauchuk, G. Bueno, M. M. Fernández-Carrobles, I. Serrano, Ó. Déniz, D. Racoceanu, and R. Venâncio. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*, 318:2199–2210, 2017.
- [20] S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. I. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Koethe, F. A. Hamprecht, and A. Kreshuk. ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*, Sept. 2019.
- [21] V. Buhrmester, D. Münch, and M. Arens. Analysis of explainers of black box deep neural networks for computer vision: A survey, 2019.
- [22] K. Chellapilla, S. Puri, and P. Simard. High performance convolutional neural networks for document processing. 10 2006.
- [23] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs, 2016.
- [24] F. Chollet. Xception: Deep learning with depthwise separable convolutions, 2017.
- [25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [26] N. Dimitriou, O. Arandjelović, and P. D. Caie. Deep learning for whole slide image analysis: An overview, 2019.
- [27] D. Dua and C. Graff. UCI machine learning repository, 2017.
- [28] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. D. Bosco, S. Walsh, D. Saltukoglu, T. L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, and O. Ronneberger. U-net – deep

- learning for cell counting, detection, and morphometry. *Nature Methods*, 16:67–70, 2019.
- [29] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.
- [30] H. Hajimirsadeghi and G. Mori. Multi-instance classification by max-margin training of cardinality-based markov networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1839–1852, 2017.
- [31] M. Halicek, M. Shahedi, J. Little, A. Chen, L. Myers, B. Sumer, and B. Fei. Head and neck cancer detection in digitized whole-slide histology using convolutional neural networks. *Scientific Reports*, 9, 10 2019.
- [32] D. Han, J. Kim, and J. Kim. Deep pyramidal residual networks, 2017.
- [33] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.
- [34] J. Hering and J. Kybic. Generalized multiple instance learning for cancer detection in digital histopathology. In A. Campilho, F. Karray, and Z. Wang, editors, *Image Analysis and Recognition*, pages 274–282, Cham, 2020. Springer International Publishing.
- [35] J. Hestness, S. Narang, N. Ardalani, G. F. Diamos, H. Jun, H. Kianinejad, M. M. A. Patwary, Y. Yang, and Y. Zhou. Deep learning scaling is predictable, empirically. *CoRR*, abs/1712.00409, 2017.
- [36] S. Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 6(2):107–116, Apr. 1998.
- [37] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks, 2018.
- [38] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.
- [39] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53(8):5455–5516, Apr 2020.
- [40] M. Khened, A. Kori, H. Rajkumar, B. Srinivasan, and G. Krishnamurthi. A generalized deep learning framework for whole-slide image segmentation and analysis, 2020.
- [41] B. Kong, X. Wang, Z. Li, Q. Song, and S. Zhang. Cancer metastasis detection via spatially structured deep network. pages 236–248, 2017.

- [42] S. Kothari, J. Phan, and M. Wang. Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade. *Journal of pathology informatics*, 4:22, 08 2013.
- [43] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, May 2017.
- [44] J. Kuen, X. Kong, G. Wang, and Y.-P. Tan. Delugenets: Deep networks with efficient and flexible cross-layer information inflows. pages 958–966, 10 2017.
- [45] N. Kumar, R. Gupta, and S. Gupta. Whole slide imaging (wsi) in pathology: Current perspectives and future directions. *Journal of Digital Imaging*, 33, 05 2020.
- [46] J. Lafferty, A. Mccallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. pages 282–289, 01 2001.
- [47] G. Larsson, M. Maire, and G. Shakhnarovich. Fractalnet: Ultra-deep neural networks without residuals, 2017.
- [48] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.
- [49] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [50] S. Lei, H. Zhang, K. Wang, and Z. Su. How training data affect the accuracy and robustness of neural networks for image classification, 2019.
- [51] G. Litjens, P. Bandi, B. Ehteshami Bejnordi, O. Geessink, M. Balkenhol, P. Bult, A. Halilovic, M. Hermsen, R. van de Loo, R. Vogels, Q. F. Manson, N. Stathonikos, A. Baidoshvili, P. van Diest, C. Wauters, M. van Dijk, and J. van der Laak. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. *GigaScience*, 7(6), 05 2018. giy065.
- [52] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafourian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis, 2017.
- [53] Z. Liu, X. Li, P. Luo, C. C. Loy, and X. Tang. Semantic image segmentation via deep parsing network, 2015.
- [54] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation, 2015.

- [55] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016.
- [56] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos. Image segmentation using deep learning: A survey. 2020.
- [57] A. M. Nguyen, A. Dosovitskiy, J. Yosinski, T. Brox, and J. Clune. Synthesizing the preferred inputs for neurons in neural networks via deep generator networks. *CoRR*, abs/1605.09304, 2016.
- [58] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation, 2015.
- [59] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.
- [60] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- [61] A. G. Schwing and R. Urtasun. Fully connected deep structured networks, 2015.
- [62] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps, 2014.
- [63] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
- [64] K. Sirinukunwattana, J. P. W. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Böhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. J. Snead, and N. M. Rajpoot. Gland segmentation in colon histology images: The glas challenge contest, 2016.
- [65] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 06 2014.
- [66] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning, 2016.
- [67] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions, 2014.

- [68] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision, 2015.
- [69] D. Tellez, G. Litjens, J. van der Laak, and F. Ciompi. Neural image compression for gigapixel histopathology image analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1–1, 2019.
- [70] H. Tizhoosh and L. Pantanowitz. Artificial intelligence and digital pathology: Challenges and opportunities. *Journal of Pathology Informatics*, 9, 2018.
- [71] J. Tomczak, M. Ilse, M. Welling, M. Jansen, H. Coleman, M. Lucas, K. de Laat, M. D. Bruin, H. Marquering, M. J. V. D. Wel, O. D. Boer, C. D. S. Heijink, and S. Meijer. Histopathological classification of precursor lesions of esophageal adenocarcinoma: A deep multiple instance learning approach. 2018.
- [72] E. Van Cutsem, R. Labianca, G. Bodoky, C. Barone, E. Aranda, B. Nordlinger, C. Topham, J. Tabernero, T. André, A. F. Sobrero, E. Mini, R. Greil, F. Di Costanzo, L. Collette, L. Cisar, X. Zhang, D. Khayat, C. Bokemeyer, A. D. Roth, and D. Cunningham. Randomized phase iii trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage iii colon cancer: Petacc-3. *Journal of Clinical Oncology*, 27(19):3117–3125, 2009. PMID: 19451425.
- [73] S. Vesal, N. Ravikumar, A. Davari, S. Ellmann, and A. Maier. Classification of breast cancer histology images using transfer learning, 2018.
- [74] M. Veta, Y. J. Heng, N. Stathonikos, B. E. Bejnordi, F. Beca, T. Wollmann, K. Rohr, M. A. Shah, D. Wang, M. Rousson, and et al. Predicting breast tumor proliferation from whole-slide images: The tupac16 challenge. *Medical Image Analysis*, 54:111–121, May 2019.
- [75] G. Wang, W. Li, S. Ourselin, and T. Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. *Lecture Notes in Computer Science*, page 178–190, 2018.
- [76] Q. Wang, Y. Ma, K. Zhao, and Y. Tian. A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, 04 2020.
- [77] X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, E. Tsougenis, Q. Huang, M. Cai, and P.-A. Heng. Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE Transactions on Cybernetics*, PP:1–13, 09 2019.
- [78] D. C. Wilbur, K. Madi, R. B. Colvin, L. M. Duncan, W. C. Faquin, J. A. Ferry, M. P. Frosch, S. L. Houser, R. L. Kradin, G. Y. Lauwers, D. N. Louis, E. J. Mark, M. Mino-Kenudson, J. Misdradi, G. P. Nielsen, M. B. Pitman, A. E. Rosenberg, R. N. Smith, A. R. Sohani, J. R.

- Stone, R. H. Tambouret, C.-L. Wu, R. H. Young, A. Zembowicz, and W. Kluetmann. Whole-Slide Imaging Digital Pathology as a Platform for Teleconsultation: A Pilot Study Using Paired Subspecialist Correlations. *Archives of Pathology and Laboratory Medicine*, 133(12):1949–1953, 12 2009.
- [79] A. M. Wright, D. Smith, B. Dhurandhar, T. Fairley, M. Scheiber-Pacht, S. Chakraborty, B. K. Gorman, D. Mody, and D. M. Coffey. Digital Slide Imaging in Cervicovaginal Cytology: A Pilot Study. *Archives of Pathology and Laboratory Medicine*, 137(5):618–624, 09 2012.
- [80] W. Yao, Z. Zeng, C. Lian, and H. Tang. Pixel-wise regression using u-net and its application on pansharpening. *Neurocomputing*, 312, 06 2018.
- [81] S. Zagoruyko and N. Komodakis. Wide residual networks, 2017.
- [82] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks, 2013.
- [83] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks, 2013.
- [84] X. Zhang, Z. Li, C. C. Loy, and D. Lin. Polynet: A pursuit of structural diversity in very deep networks, 2017.
- [85] X. Zhu, C. Vondrick, C. C. Fowlkes, and D. Ramanan. Do we need more training data? *CoRR*, abs/1503.01508, 2015.