Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Physics

Masoud Tafavvoghi

Automated segmentation of tumors in dynamic contrast-enhanced MRI of high-risk breast cancer patients undergoing neoadjuvant therapy

Master's thesis in Physics
Supervisor: Tone F. Bathen
Co-supervisor: Gabriel A. Nketiah, Neil P. Jerome, Guro F. Giskeødegård

June 2021

NTNU
Norwegian University of
Science and Technology

Masoud Tafavvoghi

Automated segmentation of tumors in dynamic contrast-enhanced MRI of high-risk breast cancer patients undergoing neoadjuvant therapy

Master's thesis in Physics
Supervisor: Tone F. Bathen
Co-supervisor: Gabriel A. Nketiah, Neil P. Jerome, Guro F. Giskeødegård
June 2021

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Physics

NTNU
Norwegian University of
Science and Technology

Abstract

Breast cancer is the most common cancer among women, with nearly 3700 new cases diagnosed annually in Norway. Within this group, between 5 and 10% are diagnosed at stage III with locally-advanced breast cancer. Presurgical tumor downstaging by neoadjuvant chemotherapy treatment makes breast-conserving surgery applicable to large tumors, and may also lower the risk of locoregional relapse.

Early evaluation of breast cancer response to treatment can show the effectiveness of that treatment, and gives the opportunity for treatment change if there is no response. Among different imaging modalities for the evaluation of tumor response, dynamic contrast-enhanced (DCE) MRI has shown the highest sensitivity in detection of the residual tumor following treatment. Typically, radiologists manually segment the tumors in each slice of the DCE image series, which is a highly time-consuming task. Moreover, manual segmentations may vary between different radiologists, leading to different estimations of the tumor volume.

The aim of this thesis was to investigate using a deep learning model for detection and segmentation of breast tumors in DCE-MR images to facilitate the measurement of tumor volume for evaluation of response to the treatments. For this purpose, a regional-based convolutional neural network (mask R-CNN), which outputs pixel-wise instance segmentation of objects, in this case tumors, was used. A dataset consisting of 111 locally-advanced breast cancer patients who underwent neoadjuvant chemotherapy in three hospitals – St Olav's University Hospital (Trondheim), Haukeland University Hospital (Bergen), and Stavanger University Hospital (Stavanger) was included in the study. The baseline and residual tumor following each treatment cycle was manually segmented. The segmentations of the St. Olav's dataset, was validated by a radiologist, hence, the model was trained on "researcher-drawn" segmentations of the images from the other institutions, and separately validated using the images from St Olav's Hospital.

The performance of the model in segmentation of the breast tumors was evaluated by sensitivity, precision, specificity, and accuracy measurements of the model. Also, to assess the model's segmentations, the dice similarity coefficients (DSC) between manual segmentations and the model predictions was calculated.

The model's accuracy in detection of the breast tumors was 0.84 with sensitivity and specificity of 0.75 and 0.71, respectively. Also, the average DSC of the test set was 0.84. Based on the achieved results, it can be concluded that the deep learning model performs well both in detection and segmentation of the breast tumors in DCE-MR images. The achieved results were quite good especially since the test cohort is completely independent from the training cohorts. However, using images with different contrast to noise ratios (CNR) in the training step could improve the performance of the model, by decreasing the

false positive detections. A future study should investigate the potential beneficial effect of adding part of the test set to training set and further evaluation of the model on the remaining cohort.

Preface

This thesis was written for the department of Physics at the Norwegian University of Science and Technology (NTNU) in spring of 2021. The subject for the thesis was defined in cooperation with my supervisor Prof. Tone F. Bathen at the Department of Circulation and Medical Imaging. I would like to thank my supervisor for her invaluable supervision, continuous support, and patience during my master's project. I am indebted also to my co-supervisors Dr. Gabriel A. Nketiah, Dr. Guro F. Giskeødegård and Dr. Neil Peter Jerome at MR Cancer Group for their technical support and kind help on my thesis.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Background and problem definition

Breast cancer is the most common type of cancer among women in Norway, with nearly 3700 new diagnosed cases annually [1]. Within this group, between 5 and 10% are diagnosed at stage III with locally-advanced breast cancer. These patients have poor prognosis and as a result lower survival rate. Presurgical tumor downstaging by medical therapy prior to surgery makes breast conserving surgery applicable also to large tumors and lowers the risk of a locoregional relapse significantly. This is currently done through medical therapy with chemotherapeutics or endocrine drugs.

Response of the breast cancer to neoadjuvant chemotherapy is one of the key factors in opting for breast conserving surgery. Also, early evaluation of the tumor response to treatments allows timely change of chemotherapy agent or treatment plan if the response is improper. There are several imaging modalities for evaluation of tumor response to treatments such as mammography, ultrasonography, and magnetic resonance imaging (MRI). Among them, MRI, especially dynamic contrast enhanced (DCE) MRI has shown the highest sensitivity in detection of the residual tumor following neoadjuvant chemotherapy [2].

Clinically, the assessment of tumor response to the treatment is done by estimating or measuring residual tumor volume after each treatment cycle. This can be done by palpation and measurements performed by calipers. However, image-based measurements are more accurate. Typically, radiologists manually segment the tumors in each slice of the image series used. However, such image series (e.g. DCE-MR images of locally-advanced breast cancer) can be voluminous since there are usually many slices containing tumor. This makes the process of manual segmentation highly time consuming. Moreover, manual segmentation is prone to intra and interobserver variabilities, leading to different estimations of the tumor volume. Recently, the use of artificial intelligence in medicine have gained attention due to its ability to facilitate and improve clinical decision-making [3]. Deep learning algorithms for instance have been found capable of making significant improvements in the accuracy of pathological diagnoses or detection and assessment of disease stages [4, 5].

1.2 Objective

The objective of this thesis is to train a deep learning model for automatic detection and segmentation of breast tumors in DCE-MRI to facilitate the measurement of tumor volume for evaluation of response to the treatments. The resulting model's output will be compared with the radiologist-validated segmentations to assess the performance of the model.

1.3 Thesis outline

The structure of the thesis is as follows: Chapter 2 contains background about breast anatomy and cancer, basic principles of MRI and its applications in breast cancer. The last part of this chapter introduces convolutional neural networks and the Mask R-CNN, which is used in this work for the detection and segmentation of the tumors in DCE-MR images. In chapter 3, datasets, methodology of the data preparation for training the model, and evaluation metrics of the model are described. Thereafter, the results are reported in chapter 4 and the evaluation of the results are discussed in the next chapter. In chapter 6, the thesis is summarized and suggestions for the future work are provided.

Chapter 2

Theory

2.1 Breast anatomy and cancer

Breast tissue consists of skin, fatty tissue, glandular tissue, connective tissue, stroma and ligaments, lymphatics, lymph nodes, and blood vessels. The glandular tissue is made up of ducts and lobules surrounded by connective tissue or stromal tissue (Figure 2.1).

The stroma are connective tissues that consist of ligaments and fatty tissues which surround the lobules and ducts. A lobule comprises approximately 30 terminal branches. Terminal ducts open into lactiferous ducts, which run towards the nipple.

Breast cancer encompasses a wide group of diseases that develop on the same basis, which is mutations in the DNA that constitutes our genes. Genes are a form of encrypted information that govern all cell functions, including the ability of cells to connect with new cells, stop dividing, and to dissemble at the appropriate time. Therefore, certain mutations in particular genes can change these instructions, leading to unrestrained and disorganized proliferation of cells and as a result the formation of a tumor.

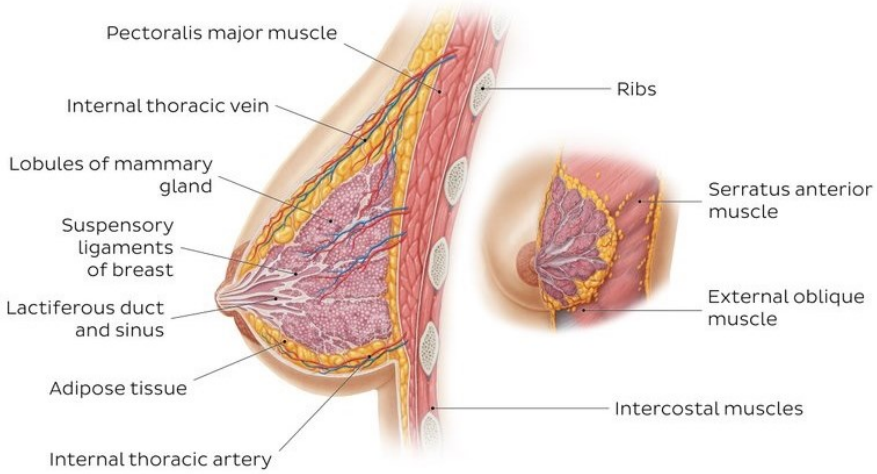


Figure 2.1: anatomy of the breast [6]. The breast contains a complex mixture of tissue types and structures, although most breast cancers arise from epithelial components.

2.1.1 Benign breast lesions

Lesions are abnormal changes in a tissue or organ that occur due to any disease or injury. Benign breast lesions form lumps and grow rapidly. Benign lesions are not invasive but enhanced proliferation of the benign lesions can increase the risk of breast cancer. They have different extent of risk and can be categorized in non-proliferative lesions, proliferative lesions without atypia, and proliferative lesions with atypia. Atypical cells are neither normal nor cancerous and they resemble healthy cells that could become cancer over time or may increase a person's risk of cancer [7]. Benign breast lesions are thought to impart no increased risk of breast cancer.

2.1.2 Malignant breast lesions

According to the type of cells and their predisposition to spread, breast cancers can be classified into in-situ breast carcinomas (ISBC) and invasive breast carcinomas. ISBC is a group of cancer cells that have not spread into surrounding breast tissue. These cells are divided into ductal carcinoma in-situ (DCIS), which appears inside the milk ducts, and lobular carcinoma in-situ (LCIS), which emerges in lobules.

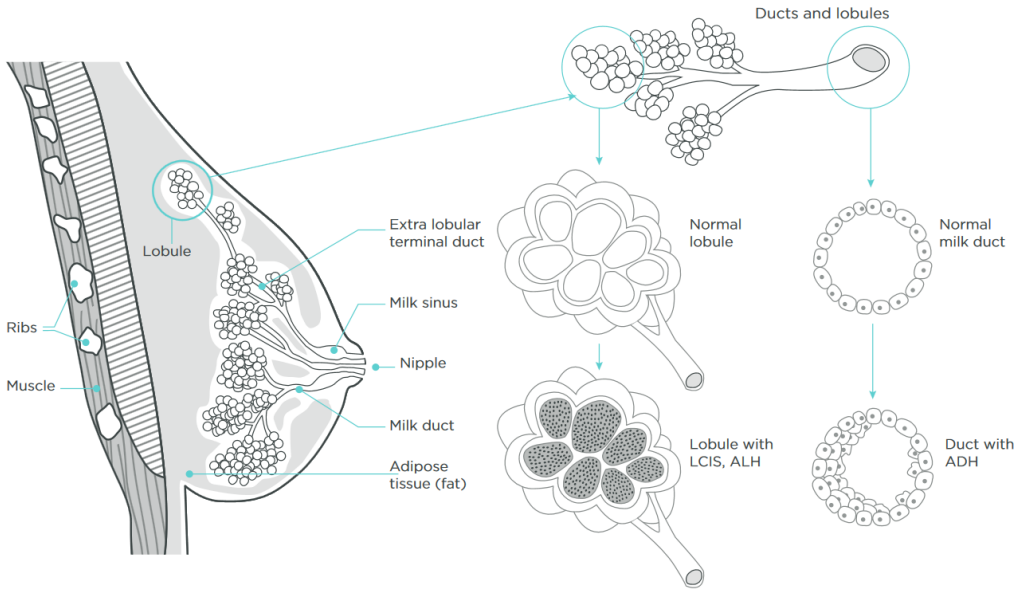


Figure 2.2: Illustration of ductal and lobular carcinoma in situ, which are both non-invasive [8]. Atypical lobular hyperplasia (ALH) is an abnormal growth of cells within breast lobules that increases the risk of breast cancer. ALH is a milder form of Lobular carcinoma in situ (LCIS) [9]. Atypical ductal hyperplasia (ADH) is a marker to an increased risk of getting breast cancer in future.

In invasive breast carcinoma, cancer cells have invaded the adjacent breast tissues outside of the duct or lobules. These cancer cells can enter the vascular system or the lymphatic system to move into other organs in the form of metastases.

Breast cancer can also be classified based on its intrinsic genetic profile, namely: HER2 (human epidermal growth factor receptor 2), luminal A, luminal B, and triple-negative. HER2-positive cancers, which overexpress the HER2 gene, create proteins that promote the cancer growth. Luminal A tumors are estrogen receptor (ER)-positive and progesterone receptor (PR)-positive and grow slowly. Estrogen

and progesterone are hormones that can attach to cancer cells and fuel the cancer growth. Cancers are called hormone receptor-positive if they have ER/PR proteins. Luminal B tumors tend to be estrogen receptor positive, progesterone receptor negative, and HER2 positive. These tumors grow more rapidly than luminal A tumors. In triple-negative tumors, the cancer cells do not contain receptors for HER2, estrogen, and progesterone. This type of breast cancer is usually invasive and usually begins in the breast ducts [10]. These more aggressive malignant tumors are the target for improvement of patients outcome.

2.1.3 Breast cancer stages

The TNM system[1] is the most common staging system for breast cancer [11]. Breast cancer staging measures the spread of the disease upon diagnosis, which is very momentous in the choice of treatment [12]. There are five stages for breast cancer – stage 0 followed by stages 1 to 4. These stages depend on the size of the tumor, site of cancer, whether cancer has spread to adjacent tissues, and whether the lymph nodes are affected or not. The characteristics of the breast cancer stages are shown in Table 2.1.

Table 2.1. Breast cancer stages [12]

Stage	Tumor size	Lymph nodes	Spreading	5-y survival rate
0	Very small, inside the glands	No cancer	confined to the breast area, not outside	100%
I	less than 2 cm	no cancer	confined to the breast area, not outside	98%
II	2-5 cm	affected by cancer	confined to the breast area, not outside	87%
III	5 cm and larger	affected by cancer; cancer has reached the muscles and skin	confined to the breast area, not outside	61%
IV	any size	affected by cancer	cancer has spread outside the breast area to any part of the body	20%.

The patients' cohort of this study includes patients with locally-advanced breast cancer (LABC) that are categorized in stages III and IV. At these stages, cancer involves the underlying muscles of the chest, skin of the breast, and/or multiple local lymph nodes. LABCs also include rapidly growing inflammatory breast cancer that makes the breast appear red and swollen.

2.1.4 Breast cancer treatments

The standard breast cancer treatment is tumor removal by surgery accompanied by adjuvant therapies, which include local irradiation and systemic therapies, such as chemotherapy and hormonal therapy. Early diagnosis and staging of the breast cancer increases the survival rates of the patients.

Breast cancer treatments can be categorized into three main groups:

[1] In the TNM system, T, N and M refer to the Tumor size, Number of adjacent lymph nodes that have cancer, and Metastasized cancer, respectively.

- Surgical excision and node dissection, in which tumor and some surrounding healthy tissues of the breast are removed.
- Radiation therapy that utilizes high energy particles or gamma rays to destroy cancer cells.
- Systemic therapy, which involves the use of medications to eliminate cancer cells. For breast cancer, four types of systemic therapies are used: chemotherapy, hormonal therapy, targeted therapy, and immunotherapy. These therapies have improved the long-term survival of the breast cancer patients significantly [13].

Neoadjuvant chemotherapy is used before the surgery to downstage locally-advanced breast cancer and potentially reduce the extent of surgery. Patients undergoing neoadjuvant chemotherapy have a higher probability of breast conservation rather than mastectomy [13]. This treatment is also most likely to be successful in a unicentric, HER2 positive, or triple-negative breast cancer and gives better survival rates and life expectancy if the response is pathologically complete [14].

2.2 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a powerful imaging modality because of its flexibility, relative safety, excellent soft tissue contrast, and high sensitivity to a broad range of tissue properties. Unlike CT and X-ray imaging which utilize ionizing radiation, MRI is a nonionizing technique. It is also noninvasive and can be used for almost any age and has a wide range of application areas including neuroimaging, angiography, pediatric imaging, musculoskeletal imaging, and cardiac imaging.

2.2.1 Basic principles of MRI

Nuclear magnetic resonance (NMR) was discovered simultaneously and independently by Bloch and Purcell in 1946 and became an important technique for chemical analyses. NMR is based on the intrinsic dipole moment of the atomic nucleus. Atoms possessing an odd number of protons and/or neutrons have a non-zero spin and thus a non-zero magnetic moment. In the presence of an external magnetic field, this magnetic moment is a combination of two states, parallel and antiparallel to the field, with different energy levels. The magnetic spins precess around the direction of the external magnetic field, with the ensemble of spins having a slightly larger fraction of the parallel state, leading to a small net magnetization vector parallel with the external magnetic field. An applied radiofrequency pulse at the correct frequency deflects the net magnetization, allowing measurement of the induced current as the net magnetic vector precesses.

Following the RF pulse, nuclei relax to their equilibrium state. There are two distinct relaxation processes, T1 (longitudinal) and T2 (transversal) that allow manipulation of the contrast in the image through choice of scanning parameters. There are many different kinds of sequences that use various combinations of repetition time, echo time, flip angle, and refocusing pulses to create T1- and T2-contrast images in a short time.

2.2.2 Breast MRI

The main indications for breast MRI are cancer staging, screening for breast cancer in women at high risk, and response evaluation to neoadjuvant chemotherapy [15]. According to the American College of

Radiology, the clinical protocol of breast MRI includes a T2 weighted scan and a T1 weighted dynamic contrast-enhanced (DCE) imaging (described below) with at least one pre-contrast and two post-contrast scans using dedicated bilateral breast coils [16]. A T2W scan, such as 2D Fast Spin Echo (FSE) sequence or 3D FSE-Cube is usually performed before the injection of a contrast agent (gadolinium-based) to confirm a benign diagnosis. T2W imaging is mainly used for qualitative evaluation of the breast anatomy, while DCE imaging provides quantitative pathophysiological information.

In DCE MRI, contrast agents are administered intravenously to change the local magnetic properties of the tissue, in this case T1 relaxation time. DCE-MRI contrast agents are paramagnetic, containing atoms (such as gadolinium, Gd) that possess at least one unpaired electron. The unpaired electrons have a magnetic moment greater than that of a proton, which reduces both T1 and T2 (depending on the concentration) of water molecules in their local environment, and as a result, increasing the signal intensity on a T1-weighted image [17]. Fast-growing or hypoxic tumors prompt growth of new blood vessels, which are often chaotic and compromised, leading to leakage of the contrast agent localised to the tumor.

Dynamic contrast-enhanced breast MRI provides excellent sensitivity to breast cancer detection and has become a useful adjunct to breast mammography. However, low specificity of MRI in diagnosing invasive breast cancer, combined with the higher cost and availability of MRI have been the limitations of this modality. Modern breast MRI stems from the introduction of gadolinium-based contrast agents, development of fast gradient-echo imaging sequences with small flip angles, and the dedicated breast surface coils in the 1980s [18].

DCE-MRI has been shown to have the highest sensitivity in the detection of invasive breast cancer (90-100%) and is independent of the density of breast tissue [19]. Experiments have shown that the concentration of the contrast agent is proportional to the relative signal increase. The signal increase relative to the image volume ($C(t)$) before the injection of the contrast agent is calculated by:

$$C(t) = (I_t - I_0) / I_0 \tag{2.1}$$

where I_t and I_0 represent the post- and pre-contrast volumes, respectively.

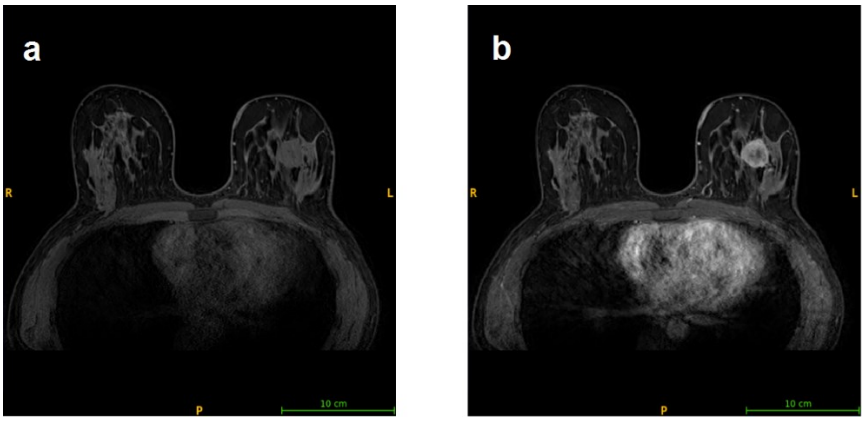


Figure 2.3: Enhancement of a breast tumor. (a) before the injection of contrast agent (pre-contrast), and (b) after the administration of contrast agent (post-contrast). (Images from QIN-Breast dataset. See section 3.1).

Lesion kinetics following the contrast agent injection should be examined carefully to increase the breast MRI specificity because cancerous lesions tend to wash-in rapidly and then wash-out, whereas benign lesions tend to enhance gradually and slowly (Figure 2.4).

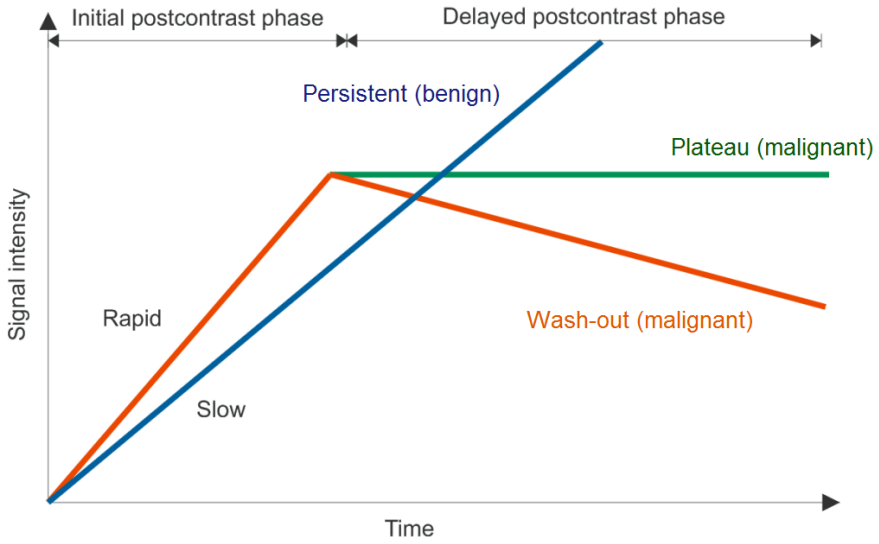


Figure 2.4: Time-signal intensity curves of benign and malignant tumors after the injection of contrast agent. (Adapted from [20]).

2.2.3 Role of MRI in locally-advanced breast cancer

Neoadjuvant chemotherapy is commonly used before the surgery to downstage locally-advanced breast cancer and to improve surgical options. It is also the standard treatment for patients with inoperable locally-advanced breast cancer [24]. Tumor response to neoadjuvant chemotherapy is one of the most important factors in opting for breast conserving surgery. Compared to other imaging modalities such as mammography and ultrasonography, DCE-MRI has the highest sensitivity in following the patients' response to neoadjuvant chemotherapy [2], and thus its usage is increasing. Chemotherapy drugs decrease tumor vasculature and as a result reduce the contrast enhancement upon imaging. The residual cancer after neoadjuvant chemotherapy often shows a better contrast enhancement on DCE-MR images which facilitates treatment response evaluation or detection of residual disease. MRI findings are also dependent on the tumor subtypes. For instance, post-neoadjuvant chemotherapy MRI for HER2 positive and hormone receptor negative tumors is highly reliable and MRI evaluations of tumor response for hormone receptor positive cancers may lead to over- or underestimation of the residual cancer [2].

Therapy response assessment by MRI can show the effectiveness of a treatment, making it possible to continue or change the treatment plan for LABC patients. An example of effective neoadjuvant chemotherapy on the breast cancer is illustrated in Figure 2.5.

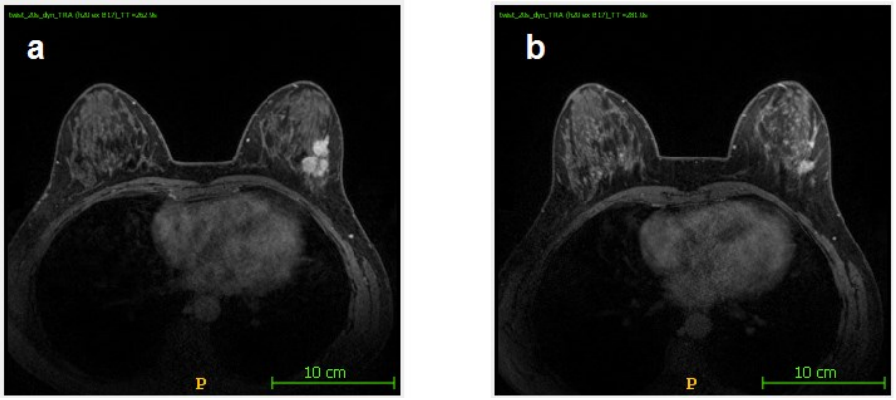


Figure 2.5: Breast tumor shrinkage due to neoadjuvant chemotherapy. (a) DCE-MR image of a breast with tumor in the left breast, (b) Same patient, one month post-neoadjuvant chemotherapy. (Images from QIN-Breast dataset. See section 3.1).

2.3 Artificial intelligence in medical imaging

Artificial intelligence (AI) has extensive applications in medical imaging today, ranging from image acquisition to image processing and analyses. Machine learning (ML) as a subfield of AI, allows computers to learn from data without being explicitly programmed. ML algorithms have become very popular in solving supervised and unsupervised classification, prediction, and recommendation tasks. Among numerous techniques of ML, deep learning is one the most promising techniques, which uses multi-layer artificial neural networks to process raw data for detection and classification problems. Deep artificial neural networks have several different architectures; among them, convolutional neural networks (CNN) have shown the highest performance in semantic segmentation problems [25]. Semantic segmentation operates by labelling each pixel of an image as a member of a class from a chosen list of possibilities, such as background, breast, or tumor, and can deliver significant information about the shape and volume of the desired class. Deep learning models with a CNN architecture have become quite powerful in recent years with comparable results performance-wise as radiologists, e.g., CT images of the kidney tumors [25]. CNN are therefore a promising tool for automated segmentation of lesions based on MR imaging, such as DCE MRI in breast cancer, with the potential to dramatically reduce the workload of a radiologist in high-risk breast cancer screening and assessment of response through neoadjuvant chemotherapy, which is the objective of this thesis.

In the following sections, the theory of the convolutional neural networks and the Mask R-CNN algorithm that is used in this thesis for detection and segmentation of the breast tumors are explained.

2.3.1 Convolutional Neural Networks (CNN)

CNNs are a subtype of artificial neural networks (ANN). ANNs were initially inspired by the human brain, which is composed of millions of neurons. These neurons/perceptrons are represented as a weighted node in an ANN, which can take and process an input, and result an output that can be used as an input for another node. These interactions enable the ANNs to be an adaptive system that can learn from data and make decision by adjusting the weights of the nodes. An ANN consists of three types of

layers: the input layer that accepts the inputs, the hidden layers for processing the inputs, and the output layer that gives the result.

Among many different architectures of the ANNs, CNNs have been shown to perform competently for computer vision tasks [26], such as object detection in self driving cars or classifications in medical images. CNNs are composed of three types of hidden layers: the convolutional layer, pooling layer and fully connected layer, of which there can be several stages depending on the architecture of the CNN. An illustration of a typical CNN architecture is shown in Figure 2.6.

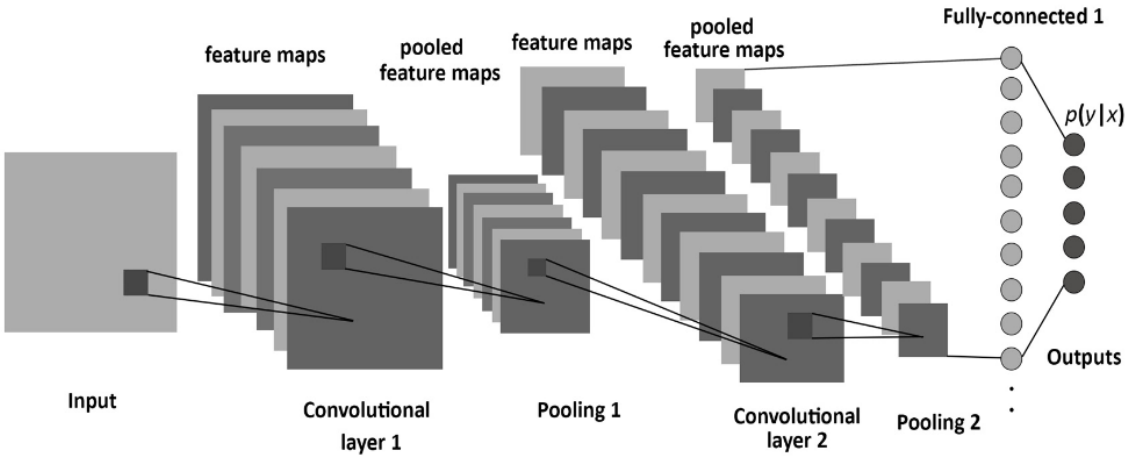


Figure 2.6: Architecture of a convolutional neural network, consisting of convolutional, pooling and fully connected layers [27]. In the case of medical images, CNNs receive the input data as a three-dimensional matrix of voxel signal intensity values at each location within the image, which makes it possible to keep the spatial relationships within the input data. At each convolutional layer, the feature maps of the input image are extracted by convolution of a specific filter with the image matrix. The feature maps are then pooled for subsampling in order to speed up the calculations and reduce the memory usage. At the end of CNN, a fully connected layer uses the generated feature maps to reach the output classes.

Convolution layer

The most important block of a CNN is the convolution layer, which extracts features (edges, colors, etc.) of the input images. To represent these features, matrices called filters (or kernels) are applied to each channel of the input image. These filters have the same dimensionality as the input image, but smaller width and height and apply a convolution operation on the input. The feature maps are then passed to an activation function, which introduces non-linearity into the output of a neuron in order to guarantee that feature map values are within a certain interval. Typically, in deeper layers of the CNN, more complex features of the image such as shape-based features can be extracted. Figures 2.7 and 2.8 illustrate the convolutional process and an example of the output matrix of this process that is called a feature map, respectively.

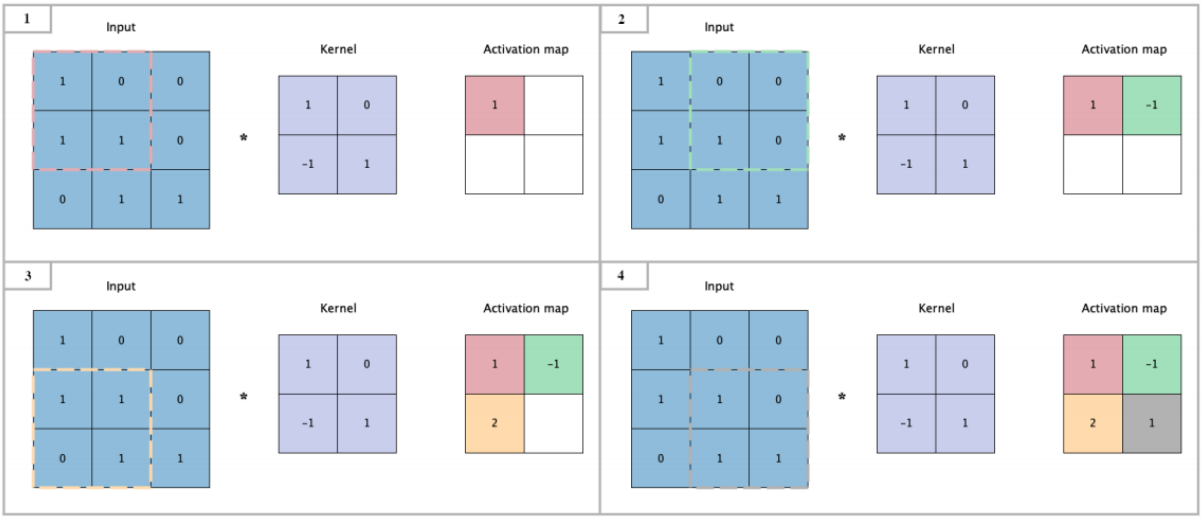


Figure 2.7: An illustration of the convolution operation. Each value in the activation map represents the result of applying the 'feature' kernel to a matching area of the input.

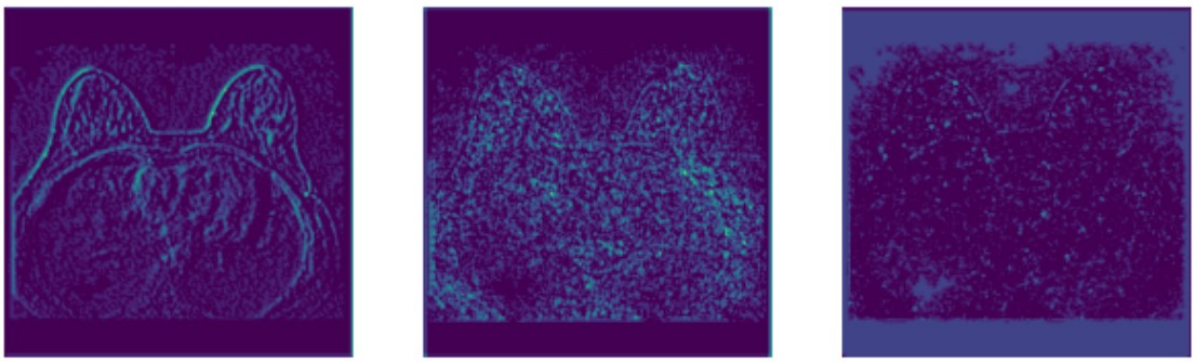


Figure 2.8: Examples of feature maps; each map is the result of convoluting the input image with a feature-specific kernel, and thus represents the degree of presence of that feature in the input image. Images are visualized by using MRCNN on public dataset.

Pooling layer

In CNN architecture, there is usually a pooling layer after a convolution layer that compresses the convolution layer output and decreases the high dimensions of the feature maps to reduce memory and computational requirements, and to prevent overfitting. Average pooling and max pooling are two of the most common methods for dimension reduction. However, max pooling can decrease the chance of vanishing gradient problem[2] which is often seen in averaging operations. The pooling layers do not have any learnable parameters, which may lead to the probable loss of valuable information.

[2] In vanishing gradient problem, the network's output changes very slightly even by a large change of parameters in earlier layers of the CNN, making the network unable to learn the parameters effectively from the earlier layers of the network.

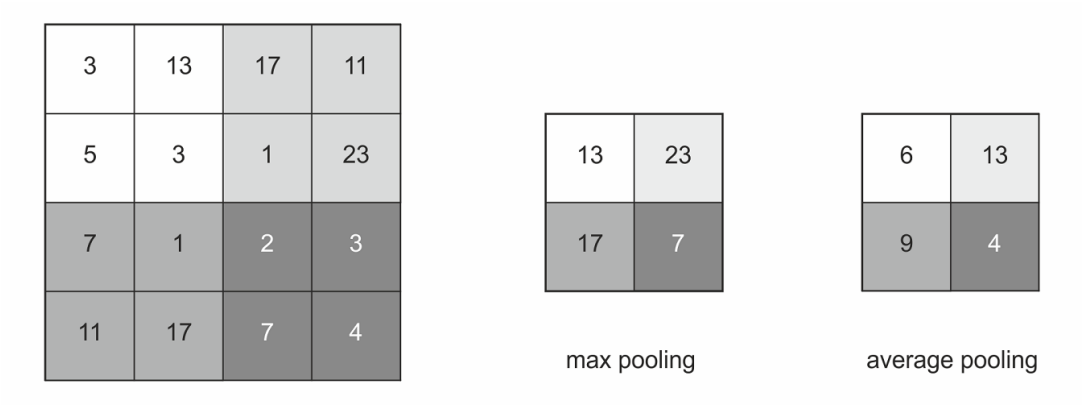


Figure 2.9: Max pooling and average pooling applied on 2x2 windows with stride 2, on this example meaning values are considered within sub-regions and only contribute to a single sub-region. Stride is a parameter of the filter that modifies the number of pixels shifts over the image. The size of the convolutional layer (left) is reduced substantially, therefore requiring reduced memory and allowing faster computation in subsequent steps.

Fully connected layer

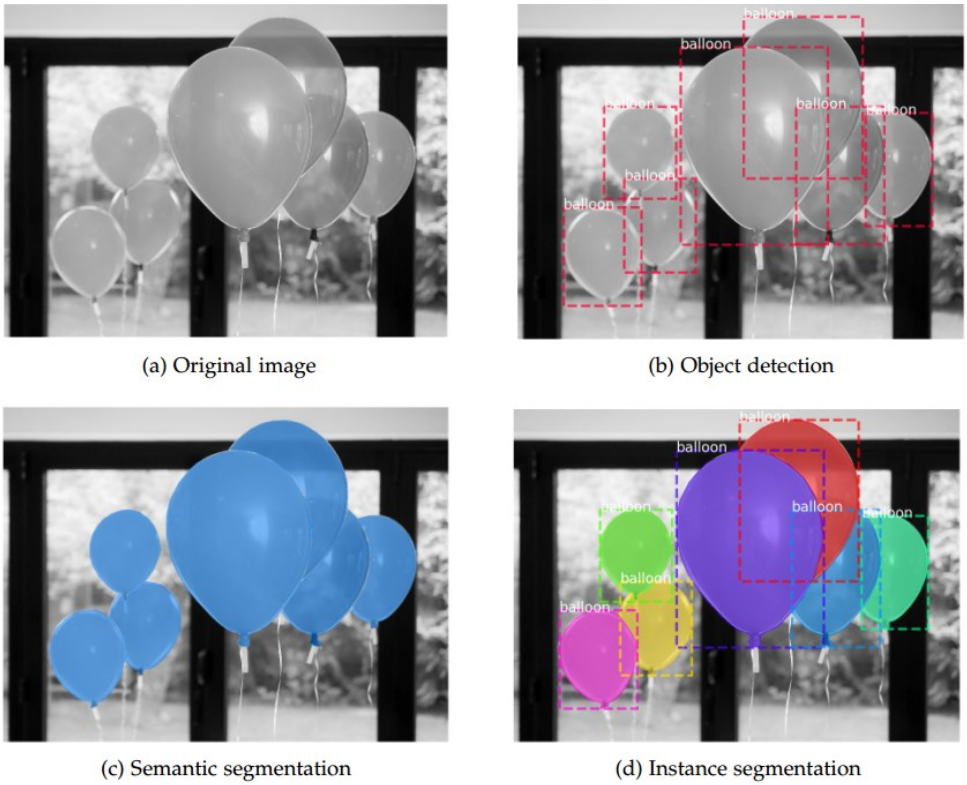
Fully connected layers use the feature maps to reach the output classes. These layers are usually added at the end of CNNs, in which all the neurons are connected to all the neurons in the next following layer (Figure 2.6). At the last layer, an activation function such as softmax[3] or sigmoid[4] function is applied to give or generate the class probability of each label (see section 2.3.4).

2.3.2 Object detection and segmentation using CNNs

Object detection is identifying objects in an image, localizing them by a bounding box and classifying them into certain classes. Image segmentation is the process of assigning the class labels to each pixel in the image. There are several different methods of image segmentation, among them semantic and instance segmentation are the most used methods in deep learning.

In semantic segmentation, pixels of the objects that belong to the same class are clustered together and given the same color/label value. Instance segmentation takes one step further and distinguishes between pixels of the same class and creates individual masks for each object in the image. An illustration of the semantic and instance segmentations is shown in Figure 2.10.

[3] Softmax function turns a vector of numbers into a vector of probabilities between 0 and 1.
[4] Sigmoid or logistic function converts the input numbers into a value between 0 and 1. It returns 0 for values smaller than 0 and 1 for values greater than 1. Therefore, output of sigmoid function can be interpreted as a probability.



(a) Original image (b) Object detection

(c) Semantic segmentation (d) Instance segmentation

Figure 2.10: Difference between semantic and instance segmentations. In semantic segmentation, all the balloons are clustered together with the same color value. On the other hand, instance segmentation creates separate masks for the objects of the same class [28].

In this work, a region-based deep neural network (Mask R-CNN) has been used for object detection and instance segmentation of the MR images. Typically, a region-based object detection framework follows three steps. First, the regions of interest (RoI) or bounding boxes are generated via an algorithm such as region proposal network (see 2.3.4). Then features maps of the each bounding-box is extracted to decide whether a bounding box contains an object or not. In the last step, by using an algorithm such as non-maximum suppression (see 2.3.3) the best bounding box, is selected and passed to the classifier to classify the detected object in the bounding box. The Mask R-CNN frameworks is discussed in greater details in section 2.3.4.

2.3.3 Important concepts in object detection and segmentation

Region of Interest (ROI)

In mask R-CNN, a region of interest (or bounding box proposal) is a tight rectangle that encloses the object of interest in an image. ROIs can be generated by various algorithms such as region proposal network (RPN) (explained in section 2.3.4) or a selective search algorithm[5]. An ROI is normally taken to mean the tumor outline, so in order to distinguish between ROIs and bounding boxes, in the following

[5] Selective Search is a fast algorithm for proposing regions in object detection tasks, based on regions' shape, texture, colour and size.

sections the ROI will be used for the tumor outlines. A bounding box is generally described by a 4x1 vector that represents the coordinates of the center of the box and the height and width of the box.

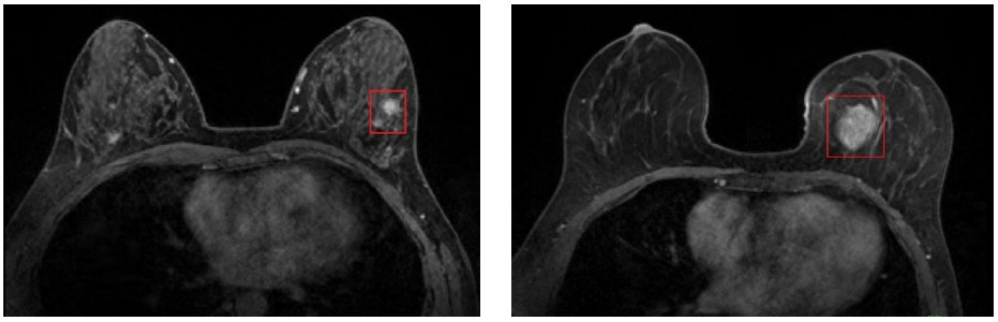


Figure 2.11: Two bounding boxes marked by red color, encompassing an object of interest (tumor).

Intersection over Union (IoU)

Intersection over union is an evaluation metric that measures the similarity between two regions, for example between model's predicted ROI and the radiologist-defined ROI which is considered the 'ground truth'. IoU ranges from 0 (no overlap) to 1 (perfect agreement), and is calculated as follows:

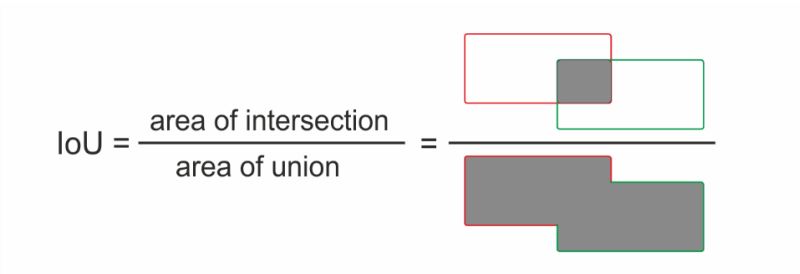


Figure 2.12: Definition of IoU

Non-Maximum Suppression (NMS)

The purpose of the NMS algorithm is to take the best bounding box out of many overlapped boxes that encompass the objects in an image by merging all the bounding boxes that belong to the same object. It takes the highest scoring detection and discards the bounding boxes that have an IoU greater than a pre-defined threshold.

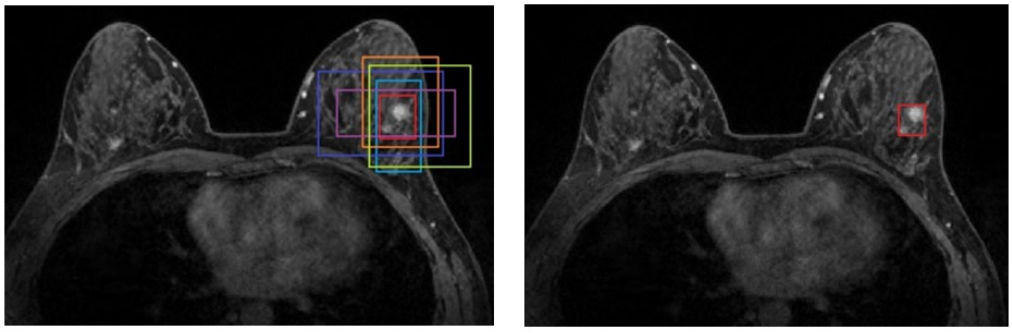


Figure 2.13: Applying the Non-maximum suppression on the proposed bounding boxes. (a) before applying the NMS, (b) after applying the NMS, only one bounding box with highest score is remained.

2.3.4 Mask R-CNN

Mask regional convolutional neural network (R-CNN) is a region-based CNN for instance segmentation tasks that can detect, classify, and segment objects in an image [29]. Being robust, easy implementation, generalizing well, and license compatibility are the main reasons for opting for Mask R-CNN in this work. Mask R-CNN is the combination of a faster R-CNN algorithm with a fully convolutional network (FCN)[6]. A Faster R-CNN object detection network is composed of a feature extraction network followed by two trainable subnetworks: a region proposal network (RPN) for generating object proposals and a network to classify the objects. The RPN is trained to generate the region proposals directly without using any external algorithm such as selective search and is located after the last convolutional layer. The framework of the Faster R-CNN is illustrated in Figure 2.14.

The RPN outputs a set of proposals, each with a probability of being an object along with objects' label. Typically, 200,000 overlapping region proposals (called anchors) are generated by the RPN for an image, that are passed to the Faster R-CNN algorithm. At this stage, anchors with the low foreground scores or those overlapping too much are removed using a non-max suppression technique. The top proposed ROIs are then passed to the next stage to be classified and to refine the bounding boxes. In order to refine the bounding boxes, a fully connected layer applies regression on the bounding boxes to best fit the detected objects.

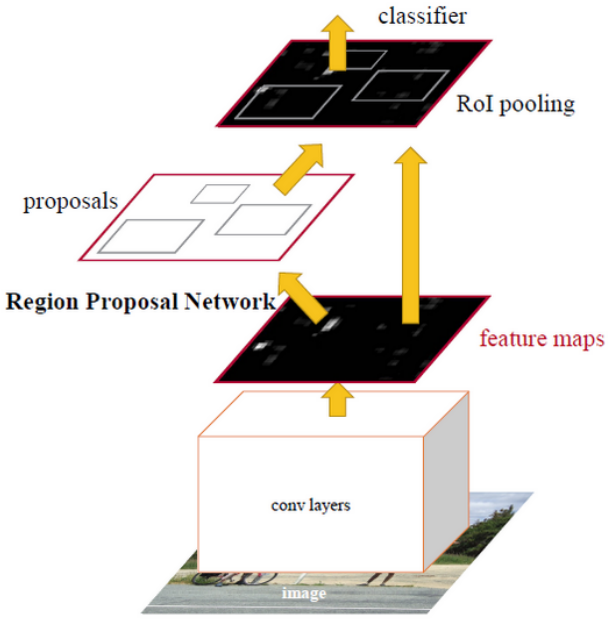


Figure 2.14: Faster R-CNN framework [30]. The RPN is inserted after the last convolutional layer of the CNN to generate the region proposals in the image. Thereafter, ROI pooling and bounding box regression are applied on the proposed regions to refine the bounding box of the objects. In the end, a classifier predicts the classes of each detected object.

[6] A fully convolutional network (FCN) contains 1x1 convolutions that perform the task of dense layers. Detailed description of FCN can be found in [34].

Thus, the faster R-CNN algorithm gives two outputs for each candidate object, a bounding box and a class label. By adding an FCN to the faster R-CNN, a binary mask of each object in the bounding boxes can be generated, which indicates pixels where the object is in the box. Mask R-CNN predicts one mask per class for every bounding box, independent of other masks. Figure 2.15 shows the framework of the Mask R-CNN.

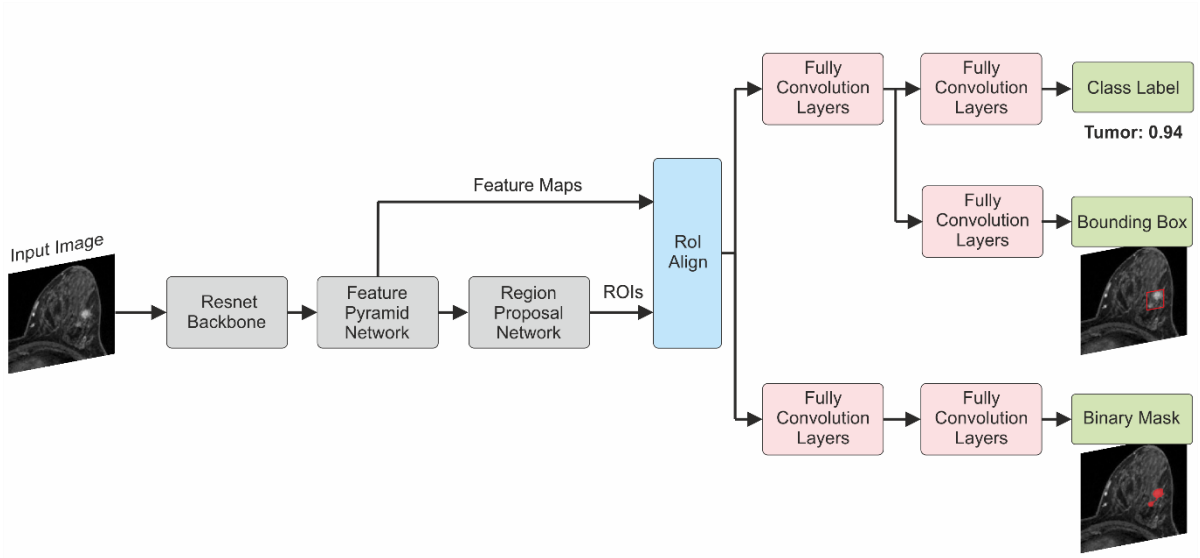


Figure 2.15: Mask R-CNN framework. Resnet 101[7] is the backbone of Mask R-CNN, which allows to train extremely deep neural networks successfully. (Interested reader can find detailed descriptions of the residual networks in [31]). The feature maps acquired from the input image and region proposal generated by the RPN are passed to the ROI align block, which extracts a small feature map from each proposed region. The refined bounding boxes, class labels, and binary masks are then generated by three independent fully convolutional network for each candidate object.

[7] Residual Network 101 (ResNet 101) is a CNN consisted of 101 layers. ResNet can handle the vanishing gradient problem by skipping some layers of the network in very deep CNNs [31].

Chapter 3

Methods

In this chapter, the datasets used, and details of the Mask R-CNN implementation are described. The first part introduces two publicly available datasets, and the PeTreMaC clinical trial dataset. Then the process of manually segmenting breast tumors in medical images and converting the generated segmentation masks to an acceptable format by the Mask R-CNN are explained. Lastly, the model's implementation and evaluation metrics used to measure the model's performance are described.

3.1 Datasets

CNNs, with lots of convolutional and pooling layers along with many filters on the layers, have an extremely large number of parameters, which makes them prone to overfitting when the dataset is small. In general, performance of CNNs in detecting objects relies on a large training data to prevent overfitting and increase the network's performance on unseen test data. Collecting such large datasets can be logistically challenging, and annotating them is highly time-consuming. Therefore, two publicly available datasets were initially used to assess the performance of the Mask R-CNN on detection and segmentation of breast tumors. Following this, the model was implemented using in-house or locally collected dataset known as Personalized Treatment in High-Risk Mammary Cancer (PeTreMaC).

3.1.1 Public datasets

For the first evaluation of the Mask R-CNN on detection and segmentation of breast tumors, two publicly available DCE-MRI datasets from the Cancer Imaging Archive were used:

- **QIN-Breast-02 dataset,** collected by Vanderbilt University Medical Center and the University of Chicago in USA for treatment assessment studies in the neoadjuvant setting. . Images in this dataset were acquired at three time points for thirteen patients, including before and during the treatment.
- The **Reference Image Database to Evaluate Therapy Response** (RIDER) dataset from the National Cancer Institute of USA, containing several imaging modalities such as CT, PET, DCE-MRI, and DWI-MRI for analyzing the response to drug or radiation therapy. The RIDER breast MRI data consists of five patients undergoing neoadjuvant chemotherapy with 8-11 days intervals.

The segmentation information of both datasets were also publicly available, therefore no manual segmentation was performed for these datasets. Sample DCE-MR images of the public datasets are shown in Figure 3.1, where the typical leakage of contrast agent in the breast tumors can be seen. More information of these datasets can be found in Table 3.1.

3.1.2 PeTreMaC dataset

The **Personalized Treatment of High-Risk Mammary Cancer** (PeTreMaC) dataset comprises 111 patients with locally-advanced breast cancer, each having between two and six visits for MR imaging during their chemotherapy treatments. The study and use of the data was approved by the Regional Committee for medical and research ethics, Western Norway (identifier: 2015/1493). The data was collected from participating hospitals in Trondheim (St Olavs University Hospital), Bergen (Haukeland University Hospital), and Stavanger (Stavanger University Hospital). These data were acquired with different MR protocol parameters and scanners (1.5 and 3T), although all contained DCE-MRI. Figure 3.2 shows some sample images of the PeTreMaC dataset, in which the effect of variation in protocols on specific contrast can be seen. More specific details of the PeTreMaC dataset including dimensions of the DCE image sets used for this work are summarized in Table 3.1.

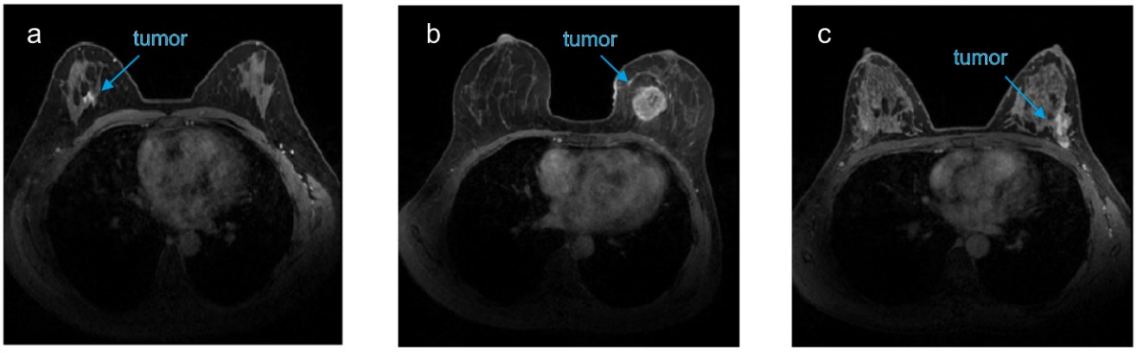


Figure 3.1: DCE-MR images from the public datasets showing the leakage of contrast agent in breast tumors. (a) Small tumor in the right breast of the patient. (b) A large round shaped tumor in the left breast of the patient. (c) Patient with tumor in her left breast close to the chest cavity.

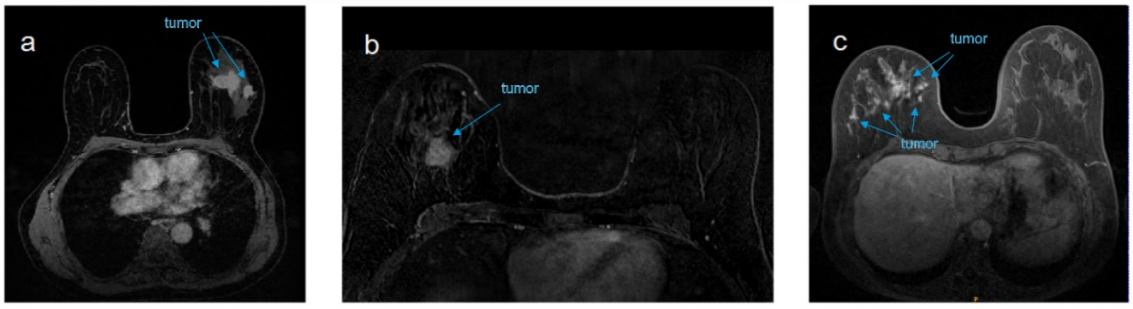


Figure 3.2: DCE-MR images of the PeTreMaC dataset, illustrating the different resolution and contrast to noise ratios (CNR) across the dataset. (a) Sample image from the Bergen dataset showing tumor in the left breast of the patient. (b) Sample image from the Trondheim dataset, with 2:1 aspect ratio and lower CNR compared to other datasets, showing a tumor in the right breast of the patient. (c) Sample image from the Stavanger dataset with tumor in the right breast of the patient. Note that both Bergen and Stavanger datasets have images with aspect ratios of 1:1 with full transverse coverage of the patient.

Table 3.1. Characteristics of the PeTreMaC and public datasets

	Trondheim	Bergen	Stavanger	QIN & RIDER
No. of patients	26	53	32	18
Total no. of visits	67	189	102	49
Image size [width, height, #slices]	512 x 256 x 120	448 x 448 x 160 416 x 416 x 144	512 x 512 x 88 528 x 528 x 170	320 x 320 x 120 288 x 288 x 60
Total DCE images	8040	28656	10010	4920
Tumor-containing images	736	3204	934	472
Imaging plane	transverse	transverse	transverse	transverse & sagittal
Contrast agent	Dotarem	anonymized	anonymized	Prohance
Slice thickness	2.5 mm	0.9, 1.1 mm	1.1, 2 mm	1.4, 2.5 mm
Magnetic field strength	3T	1.5, 3T	1.5, 3T	3T

3.2 Manual segmentation

Manual segmentation of the tumors provides the reference classifications of the image voxels from which the CNN learns to perform the segmentation, and so this is a critical first step that sets the target for the model to learn. For the manual segmentation and annotation of the DCE-MR images in this work, ITK-SNAP software was utilized, which can read and visualize diverse medical imaging formats, and segmentation files. To do the segmentation process, a polygon shape around the tumor area is drawn by hand and its corresponding mask is generated by the software. The number of slices that should be segmented for each DCE-MR volume (each patient visit) depends on the size of the tumor and thickness of the slices; in general, for locally-advanced breast cancers such as in the PeTreMaC dataset, the breasts are fully covered by the scan volume, while only a fraction will contain a tumor and require segmentation. In case of the PeTreMaC dataset, 5018 images were segmented and annotated manually. An illustration of the ITK-SNAP environment can be seen in Figure 3.3.

For use in clinical decisions and management, tumor segmentations need to be either drawn or validated by a radiologist, which essentially establishes a 'gold standard' description of the tumors. Part of the PeTreMaC dataset (patients in Trondheim) was already segmented and validated, but the remaining images from Bergen and Stavanger, comprising almost 85% of the dataset, were not. It was decided to segment (by research student M.T) all the PeTreMaC MR images specifically for this work, which would allow use of the full dataset with consistent (if non-validated) segmentations for assessment of the machine learning performance. This would also provide the opportunity to assess how much similarity there was between the researcher-drawn segmentations and the radiologist-validated ones, in order to indicate the reliability of the model to provide segmentations comparable to those validated by a radiologist.

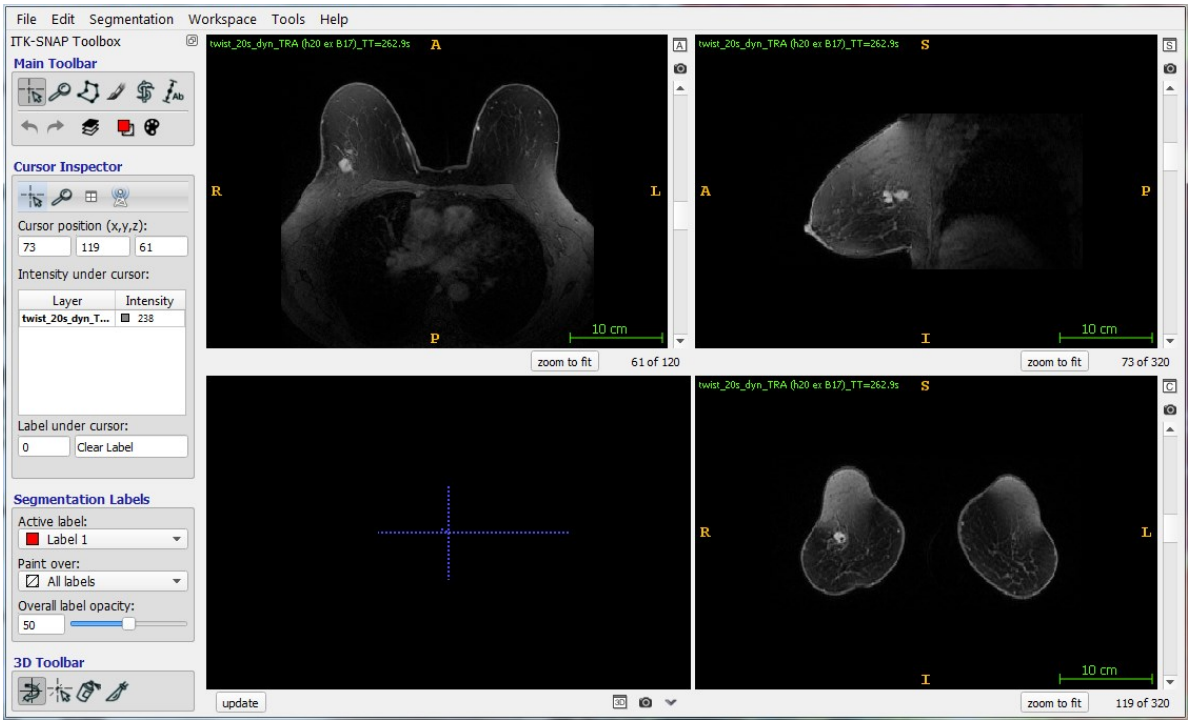


Figure 3.3. ITK-SNAP environment for visualization and segmentation of the medical images. The top-left window shows the transverse image of the breast acquired by the MR machine. Top- and bottom-right windows show the sagittal and coronal views of the breast, which are reconstructed by the software based on the transverse image. Also on the left, the provided tools for magnification, segmentation and annotation of the ROIs can be seen.

To accomplish the similarity evaluation of the segmentations in the Trondheim dataset, the Dice Similarity Coefficient (DSC) was calculated in MATLAB (Natick, USA) using the following:

$$DSC = 2 * |X \cap Y| / (|X| + |Y|) \quad (3.1)$$

where X and Y are the areas of radiologist-validated and researcher-drawn segmentations, respectively.

An important part of the manual segmentation is choosing the right images in the DCE time series. The current clinical DCE protocol for breast contains 8 image volumes in time (1 minute per volume), the first volume acquired before administration of contrast agent and the following 7 imaging volumes showing the contrast agent behavior in the different tissues. In healthy tissue, the agent comes through the blood vessels, but does not stay in the tissue, whereas in the lesion, the contrast agent leaks out of the vasculature into the surrounding tissue, and takes a long time to clear, giving signal enhancement in the post-contrast images. The exact behavior of the agents wash-in and wash-out gives information about the type and malignancy of the tumor. In DCE-MRI, the maximum contrast to noise ratio (CNR) for tumor is around third post-injection (4^{th} series in time) [32], and this selection was fixed for all the images. An illustration of the images CNR in time for DCE-MRI can be seen in Figure 3.4.

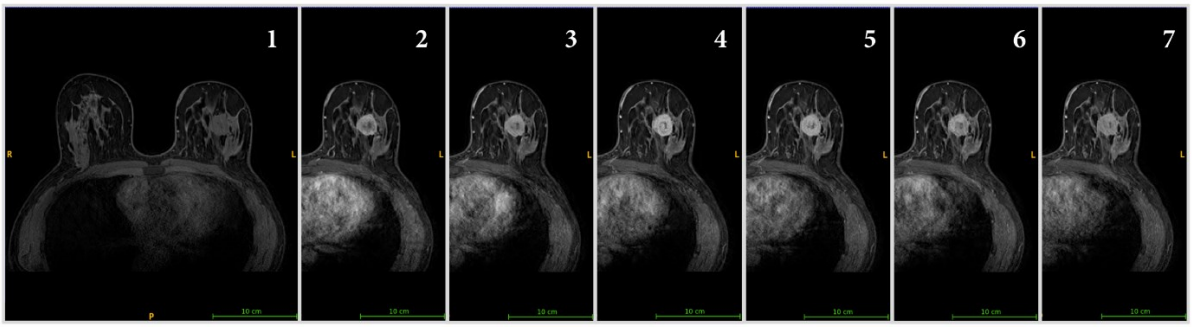


Figure 3.4: Contrast agent wash-in and wash-out in time. Contrast agent arrives after the first volume and then leaks into the tumor extravascular space, creating localized enhancement. Volume 4 was selected for use in this work as it has good tumor contrast to surrounding tissue.

3.3 Data preparation

The manually-drawn masks of the tumors are considered the 'ground truths' of the images that should be defined in the model for the training process and metric evaluations of the test set. In Mask R-CNN, to define the ground truth of an object, the pair of X and Y coordinates for all the polygon vertices of the masks are specified, and then transferred to a json metadata file, which is used as an input for the model. Figure 3.5 shows the procedure of preparing the data for training the Mask R-CNN model.

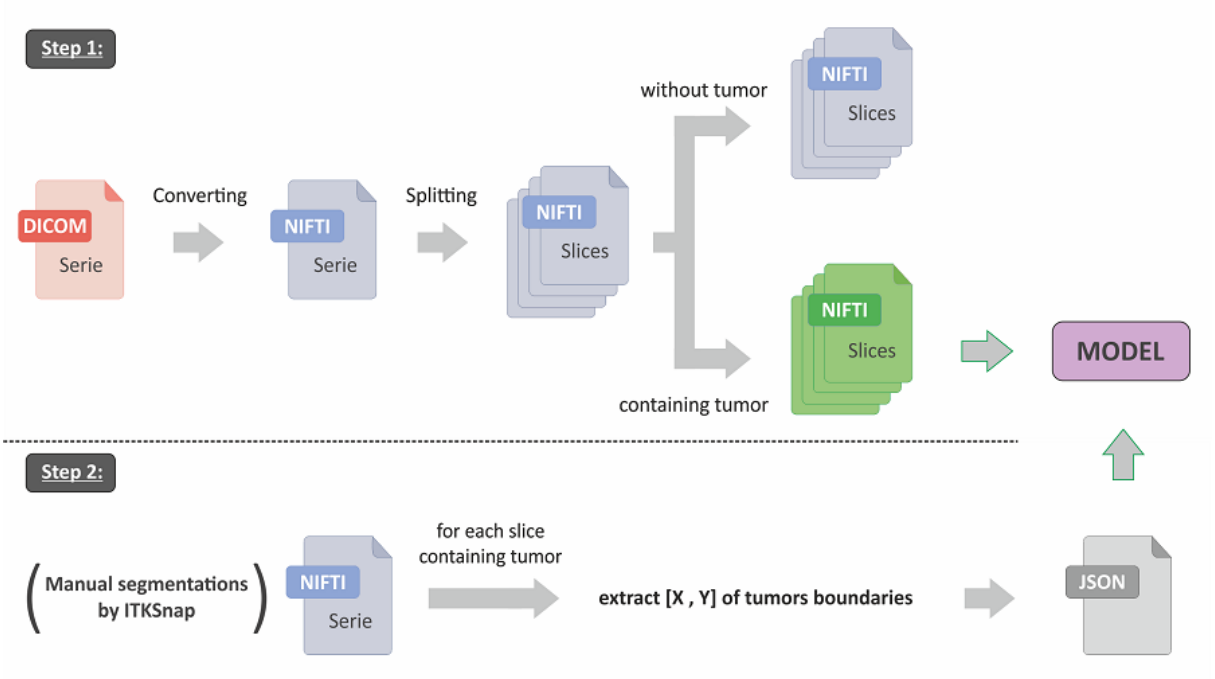


Figure 3.5: Data preparing steps for the Mask R-CNN; individual images from the patient cohort are supplied to the model along with a description of the tumor locations in the image, drawn manually. The model uses these data to 'learn' how to segment tumors in images of this type.

After the mask was created in ITK-SNAP, a custom-written Matlab code translates the segmentation to the correct json format for use in the model; Figure 3.6 illustrates this process.

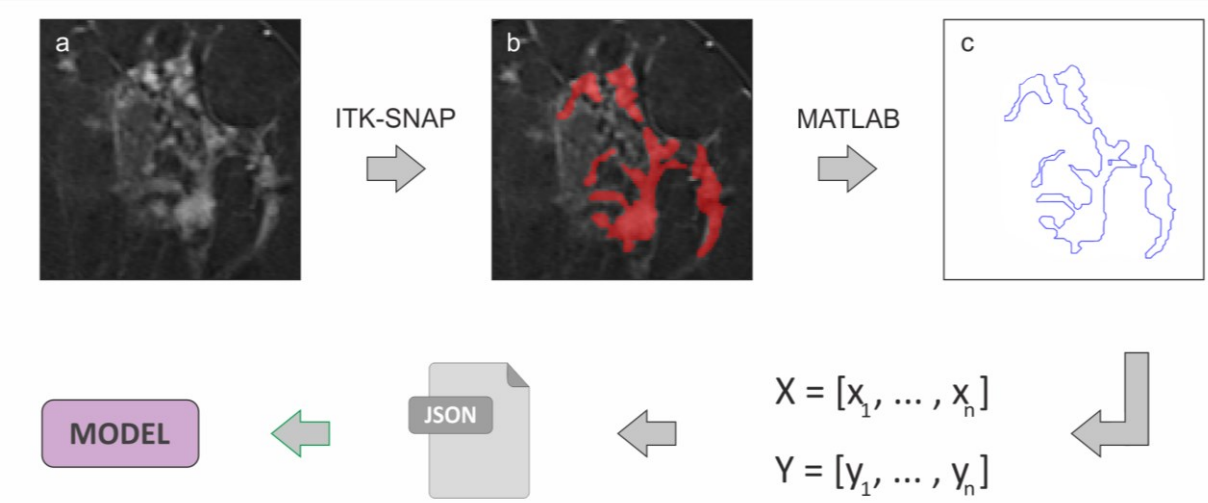


Figure 3.6: Extraction of tumor's boundary coordinates in the image, which is the second step in data preparation for training the model. (a) Magnified tumor in a DCE-MR image, (b) segmentation mask generated by ITK-SNAP, (c) extraction of [X,Y] coordinates of the tumor boundary by MATLAB.

3.4 Mask R-CNN model

The implemented Mask R-CNN code in this thesis is written in the Python language, by Matterport Inc. [28] and using libraries under the MIT License[8]. Being robust, easy to implement, generalizing well, and its license compatibility are the main reasons for implementing the Mask R-CNN. The code was cloned from the Matterport GitHub repository and altered for the specific purposes of this project. For instance, some modifications have been applied to the model for accepting gray-scale MR images that, unlike RGB images, have one channel, as well as adding functions for loading the medical imaging data and for visualization. The configuration of the model was also tailored towards the segmentation task; for example, the number of classes had to be changed to 1, as we have only 1 class of object (breast tumor) in the images. Furthermore, the number of anchors per image was decreased from 256 to 128, because in most of the images, tumors are expected to appear as a single mass (or in worse cases only a small number of separate tumor masses can be seen in an individual slice).

There is also an option in the model that allows to freeze (exclude) some layers during the training. By default, the model is set to use only the network heads, which lets the model to decrease the processing load. This functionality is suitable only if the used dataset is similar to the COCO[9] dataset since its pre-trained weights are available. The DCE-MR images look quite different from the COCO dataset, thus freezing the layers would not be appropriate in detection and segmentation of the breast cancer tumors and was disabled.

[8] MIT License grants the permission to use, copy, edit and publish a code free of charge [35].

[9] Common Objects in Context dataset is a larg scale object detection, segmentation, and captioning dataset, containing milions of everyday objects captured from everyday scenes.

3.4.1 Training and evaluation by public datasets

To train the Mask R-CNN model for detection and segmentation of the breast cancer tumors using the QIN-Breast and RIDER datasets, images were split in a 70:15:15% ratio for training, validation, and testing sets, respectively. The learning rate was set to 0.001 with 100 epochs and 100 steps per epoch, chosen as an estimate for optimal fitting.

The number of epochs determines how many times the weights of the network are allowed to change. Optimally, the number of epochs should correspond to the point where validation error is lowest. Increase in validation error with more epochs is a sign of overfitting (Figure 3.7), and training should be terminated at that epoch to take the optimal weights. Detailed explanation of this can be found in [33]. The certainty threshold of the model was set to 85%, meaning that the model will exclude detections that their probability of being a tumor is less than 85%. Configuration of this model can be found in Appendix A.

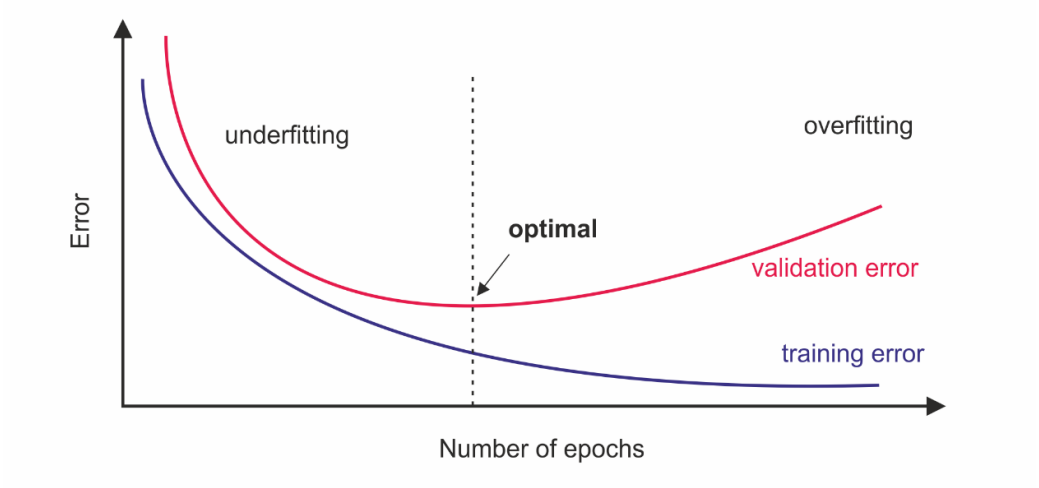


Figure 3.7: Training and validation error curves. The optimal number of epochs is where the validation error starts to increase (a sign of overfitting the model to the specific training data, leading to more error with unseen test data).

Having less than 400 images as training dataset is rather small, and as a result the model was set to initialize the training from the COCO pre-trained weights instead of starting with random weights. This alleviates the requirement for large training datasets to some degree but is limited by the overall translatability of the pre-trained weights for the new task, and allows investigation of this strategy for application of machine learning tools to smaller datasets (such as from clinical study cohorts). For this part of the work, the Google Colaboratory was used; this is a hosted Jupyter notebook service providing free access to computing resources such as graphics processing units (GPUs). The anonymized public datasets were provided to the model through Google Drive, where the Colab can access to the data.

3.4.2 Training and evaluation by PeTreMaC dataset

As described in section 3.2, part of the PeTreMaC dataset (acquired in Trondheim) was segmented and validated by a radiologist. The Mask R-CNN model was trained on researcher-drawn segmentations of the remaining data from Bergen and Stavanger, and the performance of the model was assessed by

comparing the generated masks with researcher-drawn segmentations and, ultimately, the radiologist-validated ones. Since the Trondheim dataset comprises around 15% of the overall PeTreMac dataset, with 736 slices containing tumor, the aggregate of the two other datasets were split randomly into 80:20 ratios for training and validation. Thus, training, validation, and test datasets of the PeTreMac had 68, 17, and 15% ratios, respectively. Additionally, 9 out of 291 patients visits were excluded due to very poor CNR, and low confidence in the researcher-drawn segmentation.

The PeTreMac dataset contains personal and medical information about patients and is considered as sensitive data. According to Norway's data regulations and laws, such a data cannot be downloaded and used in one's personal computer. Therefore, to work on the PeTreMaC dataset, we used the NTNU's HUNT Cloud, which gives access to many services such as GPU computing. The GPU used in this thesis, was a Tesla P100 with 16GB HBM2 memory, allowing to run the Mask R-CNN model by nearly 5000 images.

Similar to the previous experiment with the public data, the number of epochs was set to 100 but this time the learning rate was set to 0.0005 with 200 steps per epoch for training the model. This time, instead of using the COCO pre-trained weights, the output weights of the previous model (trained model by public dataset) were used to start the learning, in order to take advantage of weights that are more tailored towards the input image type. The learning rate is likely the most important hyper-parameter in configuring a CNN [34]. It controls how quickly the model should change in response to the estimated error at each epoch. A large value of learning rate can lead to sub-optimal weights in model or an unstable training process, whereas a too small learning rate may result in a long training process.

3.5 Evaluation metrics

The performance of the Mask R-CNN on detection of the tumors can be evaluated by using precision, sensitivity, and specificity concepts, three statistical measures of a binary classification test. To define these concepts, first we need to determine the four possible outcomes of a classification model. For this specific work, a positive detection is defined as the model returning a tumor region, which is either classified as a *true positive* (TP) is it corresponds to a manually-segmentation tumor, or as a *false positive* (FP) if there is no corresponding manual segmentation in the image. Correspondingly, a *true negative* (TN) is where there is no model-identified tumor regions, and there is no manual segmentation, and a *false negative* (FN) occurs where the model fails to return a tumor region where a manual segmentation exists for the image. In some slices a tumor may appear as multiple separate masses in the image; here, the model detects and segments them separately and in evaluation of the model they are considered as individual TP/FN detections. For example, Figure 3.8 shows a slice with two separate masses of tumor in which the model has detected both of them correctly. Therefore, it is counted as 2 TP detections. It is important to note that these measures refer to the tumors as a whole; the quality of the model segmentations are assessed by use of the Dice Similarity Coefficient, described earlier.

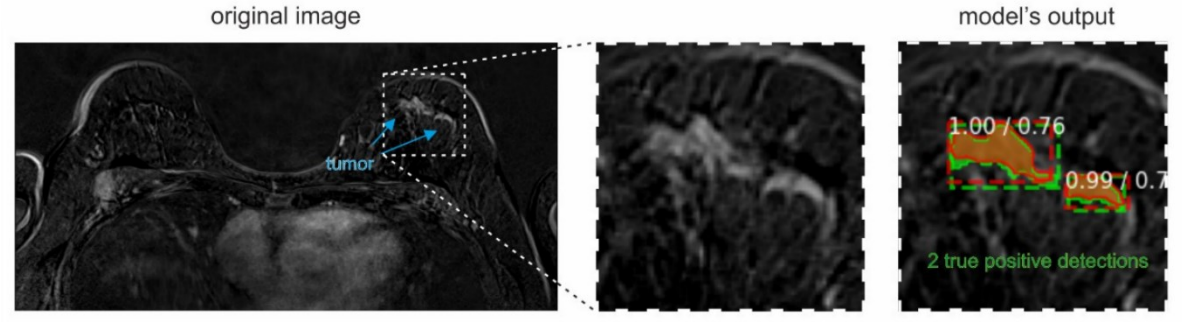


Fig 3.8: Two true positive detections of the model. The image on the left shows 2 separate masses in the left breast of the patient. On the right, the model's output is illustrated, showing 2 TP detections and their corresponding segmentations.

Precision measures the ability of a model in identifying the correct positive cases from all the positive cases and is given by

$$\text{Precision} = TP/ (TP+FP) \tag{3.1}$$

Sensitivity also referred to as recall measures the true positive rate and is given by equation 4.2. Sensitivity indicates the ability of the model in detection of all relevant cases.

$$\text{Sensitivity} = TP/ (TP+FN) \tag{3.2}$$

and specificity measures the true negative rate and indicates how well the negative class is predicted by the model. Specificity becomes more particularly important in medical applications, as the surgeons need to know what not to remove too.

$$\text{Specificity} = TN/ (FP+TN) \tag{3.3}$$

In case of our study, precision is the percentage of correctly detected breast cancer tumors by the model and sensitivity is the percentage of the detected tumors in relation to all tumors present in the MR image. For the object detection models such as Mask R-CNN, declaration of the true and false positives is dependent on the threshold of detection scores defined in the model's configuration. Thus, changing the threshold, affects precision, sensitivity, and specificity measurements.

The accuracy of the model is the other evaluation metric that is used in object detection tasks. Accuracy is the proportion of true detections among the total number of detections and is given by

$$\text{Accuracy} = (TP+TN)/ (TP+TN+FP+FN) \tag{3.4}$$

Chapter 4

Results

The final results of the Mask R-CNN model are presented in the following chapter, including the ability of the model to return tumor regions corresponding to the manual segmentations, and the similarity measures of these proposed tumor segmentations to the 'ground truth'.

4.1 Comparison of manual segmentations in the PeTreMaC dataset

Since radiologist-validated segmentations were not available for the whole PeTreMaC cohort, it was decided to use researcher-drawn manual segmentations and assess their use as a proxy. Figure 4.1 shows some examples of differences between the researcher-drawn segmentations and the corresponding radiologist-validated segmentations from the Trondheim dataset. The average DSCs comparing researcher-drawn segmentations to those validated by a radiologist can be found in Table 4.1; values represent the average DSCs over the volume for all scans, for each patient in the Trondheim dataset. The average DSC over all patients was 0.85 with a range of 0.93 to 0.65.

The DSCs are calculated for all the slices that contain tumor. In a number of cases (5% of slices), the radiologist segmentation did not contain voxels in the first/last slices which were included in the researcher-drawn segmentation. This indicates that for those cases, the radiologist volume was slightly smaller. The DSC in these individual slices is therefore equal to zero, however their low weights in calculation of the average DSC does not significantly affect the overall DSC between researcher-drawn and radiologist-validated segmentations. An illustration of slices that are not segmented by the radiologist are shown in Figure 4.2.

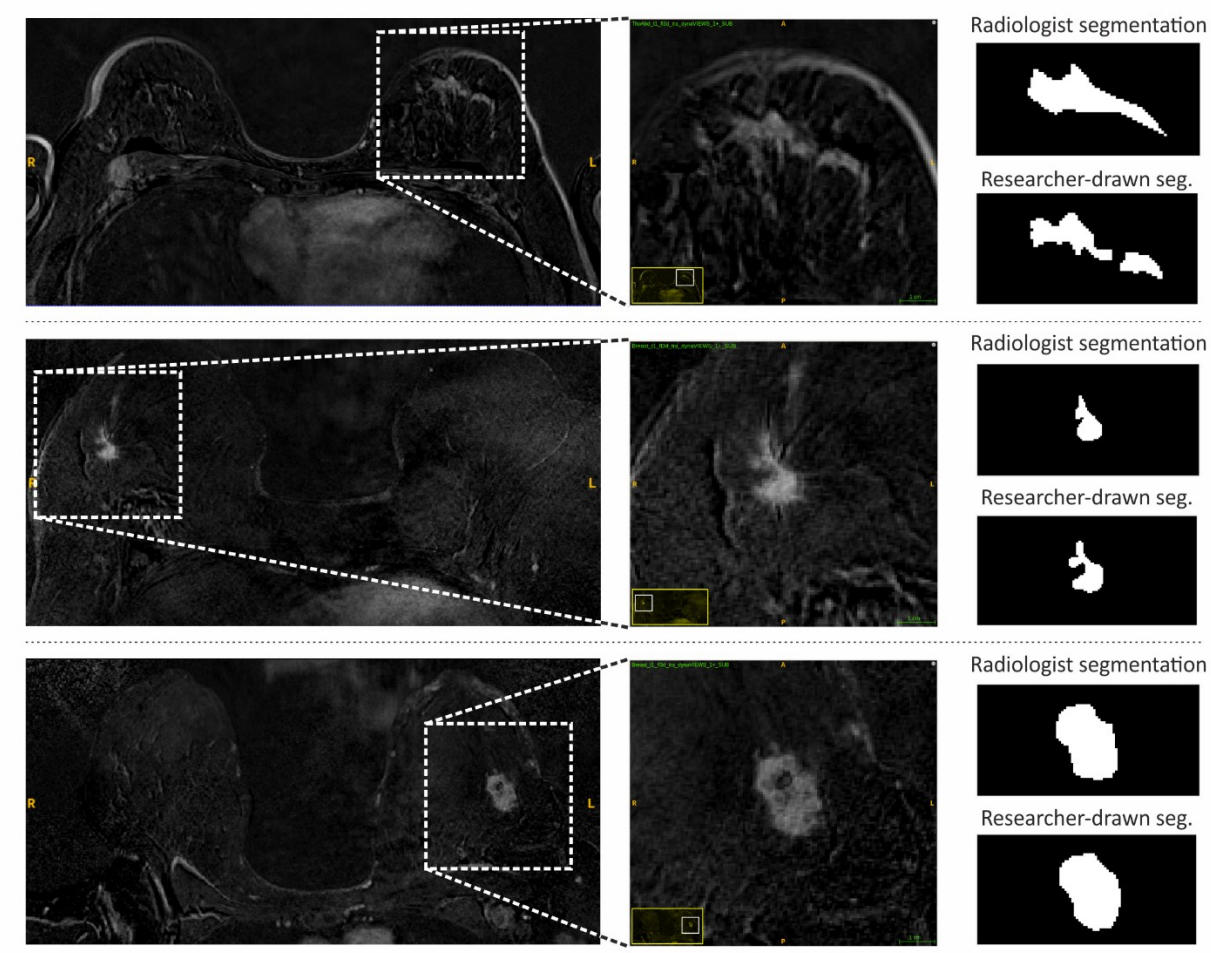


Figure 4.1: Comparison of researcher-drawn and radiologist-validated segmentations in three different patients. In general, DSC values were high, showing good agreement.

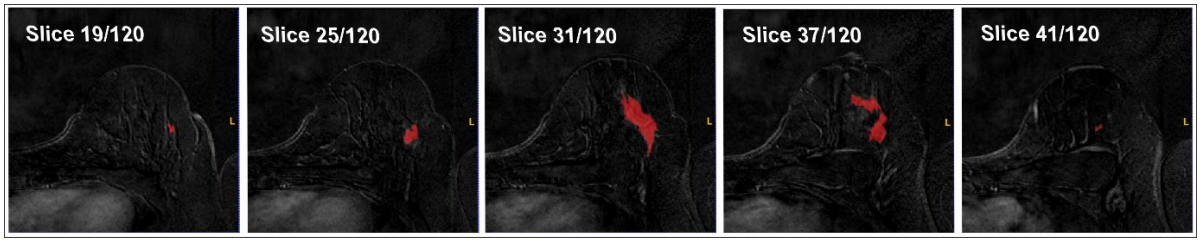


Figure 4.2: Area of segmented tumor in different slices of the same tumor volume. First and last tumor-containing slices (19 and 41) have small areas of the tumor that are not segmented by the radiologist for this patient. Researcher-drawn segmentations are highlighted in red in the images.

Table 4.1: Dice similarity coefficients between researcher-drawn and radiologist-validated segmentations of the Trondheim dataset, reported as a mean value for each patient across all scans.

Patient's ID	Number of visits	Number of segmented slices	Average DSC
PeTreMac_5001	1	10	0.93
PeTreMac_5002	4	15	0.82
PeTreMac_5003	2	17	0.88
PeTreMac_5006	4	*	*
PeTreMac_5008	1	*	*
PeTreMac_5013	4	41	0.77
PeTreMac_5014	5	83	0.77
PeTreMac_5016	2	20	0.91
PeTreMac_5017	5	31	0.84
PeTreMac_5018	2	36	0.86
PeTreMac_5019	2	11	0.76
PeTreMac_5020	2	13	0.65
PeTreMac_5021	2	18	0.90
PeTreMac_5022	3	7	0.93
PeTreMac_5024	1	8	0.88
PeTreMac_5025	3	20	0.79
PeTreMac_5026	2	9	0.89
PeTreMac_5027	2	20	0.92
PeTreMac_5028	2	44	0.84
PeTreMac_5029	3	15	0.92
PeTreMac_5030	2	20	0.77
PeTreMac_5031	3	40	0.91
PeTreMac_5032	3	47	0.92
PeTreMac_5033	2	19	0.88
PeTreMac_5034	3	43	0.84
PeTreMac_5035	2	10	0.91

* No segmentation provided by the radiologist.

Some examples of excellent and poor dice similarity coefficients are illustrated in Figures 4.3 and 4.4, respectively. As we can see, higher DSCs belong to slices with high contrast and tumors with a round/oval shape. In contrast, tumors with irregular shapes and slices containing low CNR have the lowest DSCs.

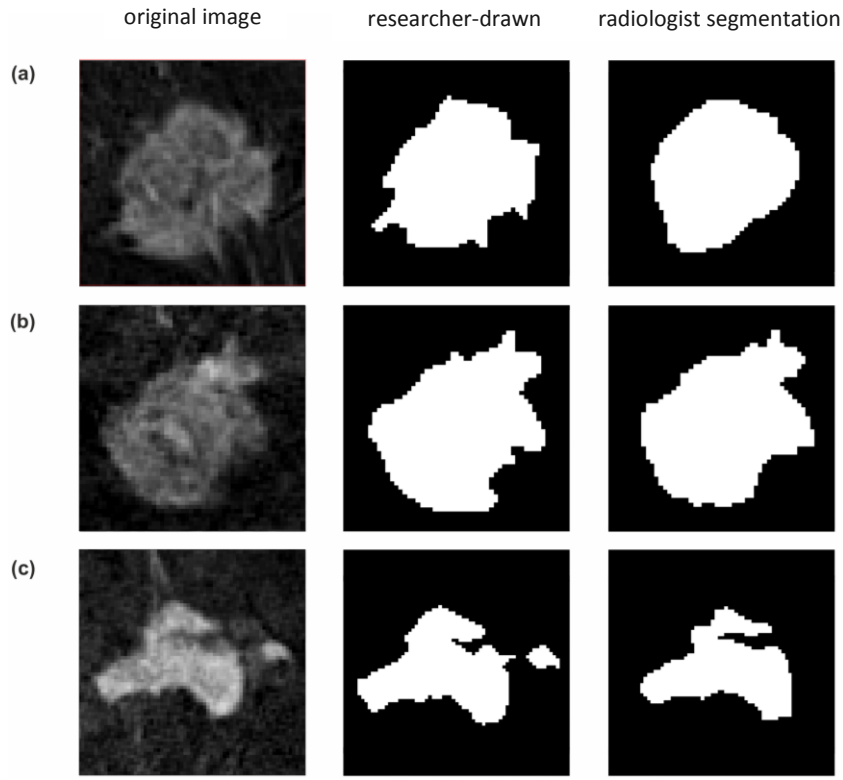


Figure 4.3: Examples of excellent DSCs between researcher-drawn (unvalidated) and radiologist-validated segmentations in Trondheim dataset. (a) Slices of tumor regions in sample patients with DSCs of 0.95 (a), 0.96 (b) and 0.92 (c), respectively.

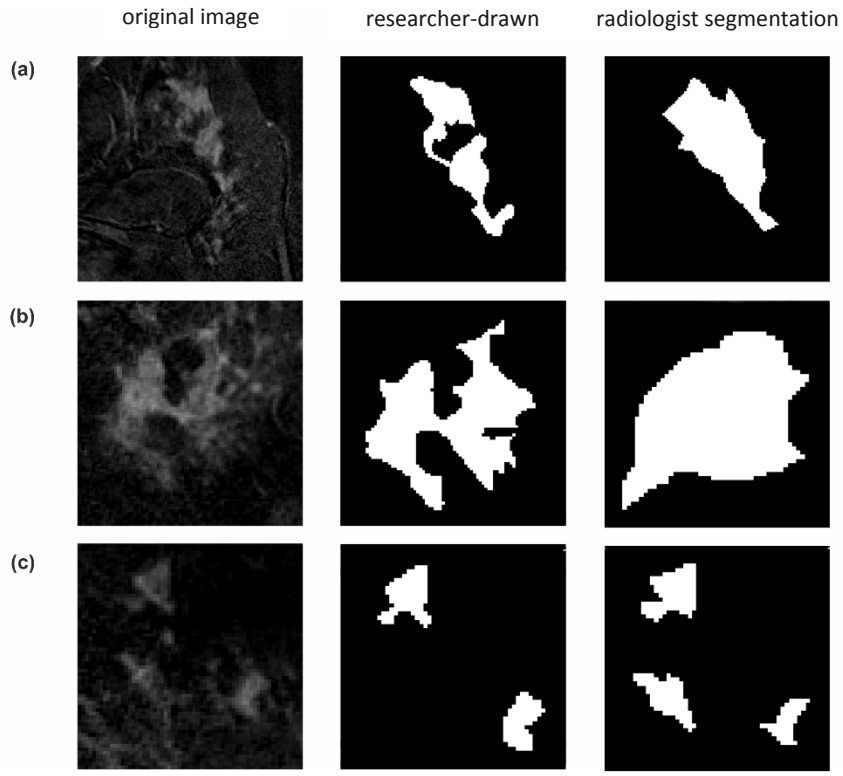


Figure 4.4: Examples of poor DSCs between researcher-drawn and radiologist-validated segmentations in Trondheim dataset. (a) Sample slices showing tumor regions in patients with DSC = 0.72 (a), 0.64 (b) and 0.64 (c), respectively.

These average DSC values illustrate for the Trondheim dataset that a researcher-drawn segmentation is sufficiently representative of a radiologist-validated segmentation for use in the machine learning model, allowing the use of all images from the cohort including those patients scanned in Bergen and Stavanger.

4.2 Evaluation of the Mask R-CNN model

4.2.1 Public dataset

The average dice similarity coefficient between the model predicted segmentations and the manual segmentations of the tumors was 0.77 for the test dataset. Also, the precision, sensitivity and specificity measurements were calculated as 0.79, 0.95 and 0.84, respectively and the accuracy of the model was 0.87. Figure 4.5 shows some examples of the model predictions with true positive, false positive, and false negative detections.

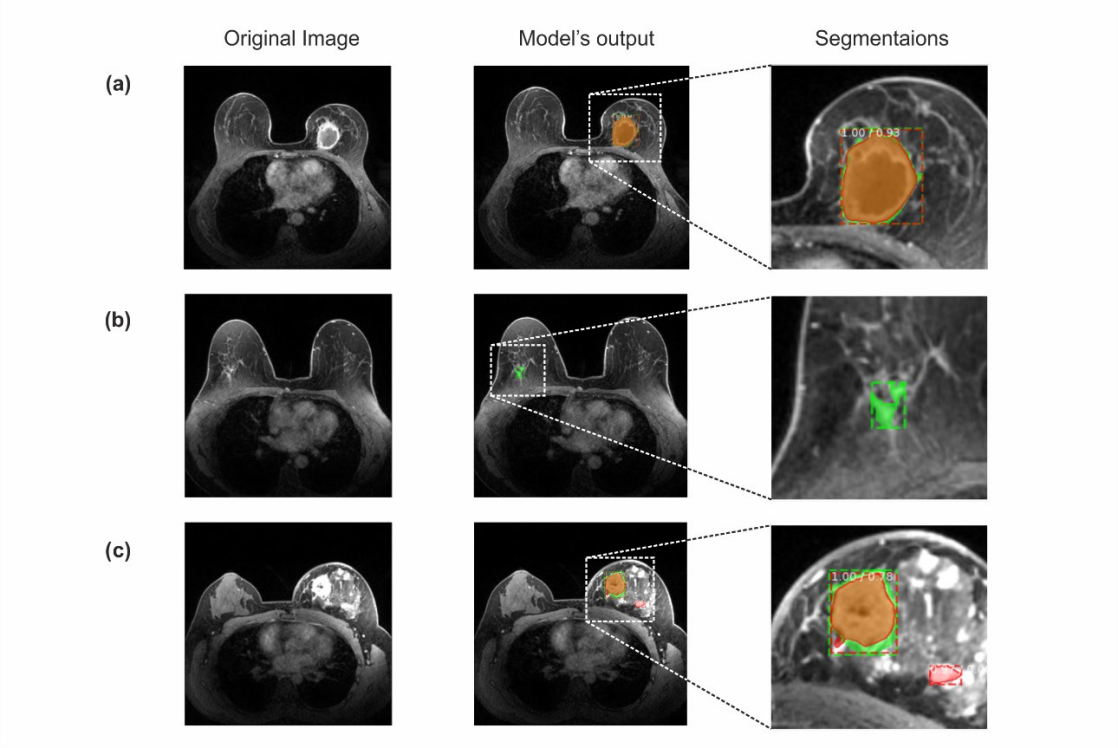


Figure 4.5: Mask R-CNN predictions. (a) True positive detection, (b) False negative detection, and (c) True positive and false positive detections. Segmentations highlighted in green and red are manual segmentations (ground truth) and model predictions, respectively. The overlap of these two masks appears orange/brown.

4.2.2 PeTreMaC dataset

Test set

Since the test dataset (Trondheim MR images) is segmented once locally and once by a radiologist, the average dice similarity coefficients are calculated for both of them to investigate which of the segmentations are closer to the model's outputs.

Figure 4.6 shows the distribution of average DSCs for comparison of manual (researcher-drawn and radiologist-validated) versus model segmentations of the test dataset. As we can see in the Figure, although the distribution of researcher-drawn segmentations has a larger range than the validated ones, median of the DSCs for the researcher-drawn segmentations is higher, which shows that the model segmentations are closer to the researcher-drawn segmentations. There is also an outlier value in the researcher-drawn segmentation with DSC of 0.57, which is the minimum DSC among all patients.

The average (mean) DSC across tumor volumes between the radiologist's segmentations and model predictions was 0.83, which is very close to the average DSC of the researcher-drawn segmentations at 0.84. In Figure 4.7 examples of the differences between these three segmentations for identical images are shown. Precision, specificity, sensitivity, and accuracy of the model were 0.75, 0.71, 0.97 and 0.84, respectively. The model's average DSCs for similarity to researcher-drawn and radiologist-validated segmentations can be seen in Table 4.2.

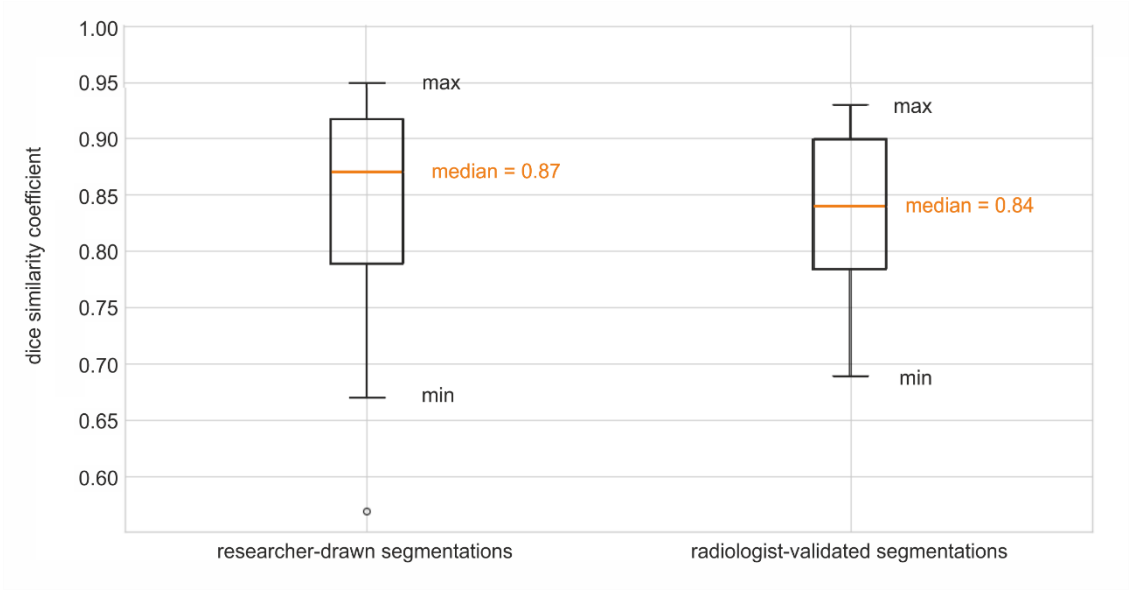


Figure 4.6: Distribution of dice similarity coefficients in the test dataset. Radiologist-validated segmentations have a smaller range of distribution with minimum and maximum DSC of 0.69 and 0.93, and median of 0.84. Minimum and maximum DSC of the researcher-drawn segmentations are 0.67 and 0.95, respectively with a median of 0.87, which is slightly greater than median of validated segmentations.

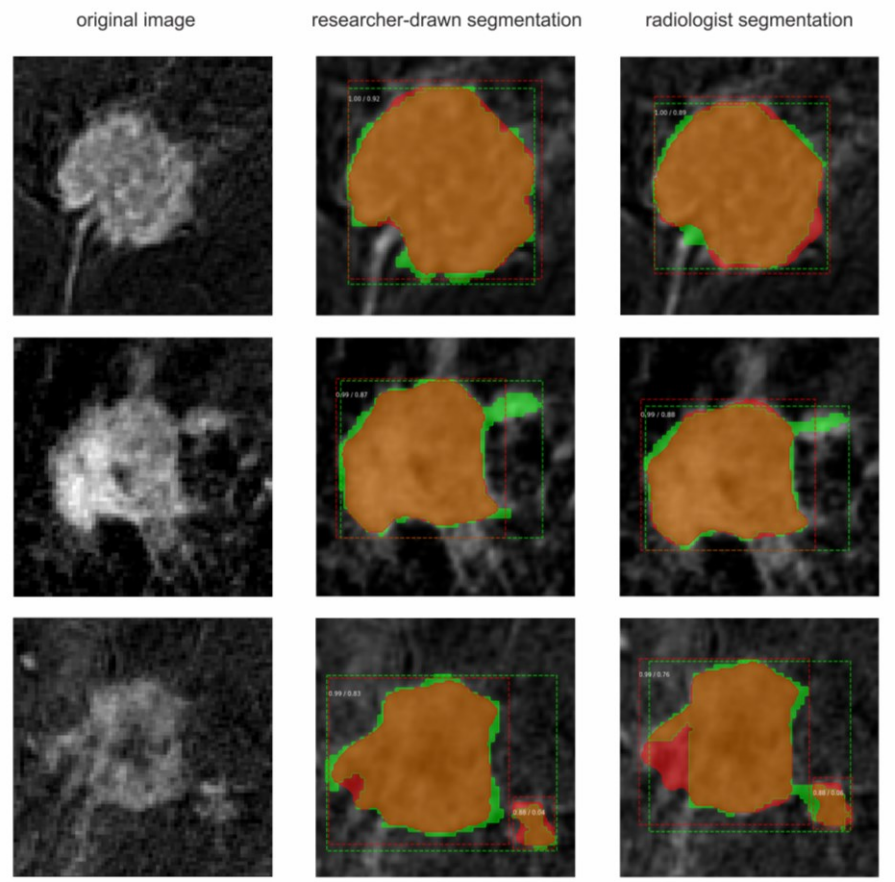


Figure 4.7: Differences between radiologist, researcher, and model segmentations for identical images. Researcher and model segmentations are highlighted in green and red colors; and intersection of them looks brown(/orange). The areas in pure green are areas of tumor that have not been segmented by the model and areas in pure red are areas that have not been segmented manually but considered as tumor by the model.

Table 4.2: Dice similarity coefficients of the Mask R-CNN segmentations

Patient's ID	Dice similarity coefficient over volume of the tumor	
	Researcher-drawn segmentations	Radiologist segmentations
PeTreMac_5001	0.95	0.93
PeTreMac_5002	0.88	0.79
PeTreMac_5003	0.92	0.87
PeTreMac_5006	0.57	*
PeTreMac_5008	0.88	*
PeTreMac_5013	0.72	0.73
PeTreMac_5014	0.71	0.69
PeTreMac_5016	0.91	0.91
PeTreMac_5017	0.80	0.83
PeTreMac_5018	0.78	0.82
PeTreMac_5019	0.79	0.75
PeTreMac_5020	0.84	0.77
PeTreMac_5021	0.91	0.91
PeTreMac_5022	0.93	0.92
PeTreMac_5024	0.88	0.85
PeTreMac_5025	0.77	0.84
PeTreMac_5026	0.92	0.88
PeTreMac_5027	0.92	0.91
PeTreMac_5028	0.80	0.79
PeTreMac_5029	0.92	0.90
PeTreMac_5030	0.67	0.74
PeTreMac_5031	0.91	0.90
PeTreMac_5032	0.86	0.80
PeTreMac_5033	0.85	0.84
PeTreMac_5034	0.79	0.77
PeTreMac_5035	0.94	0.87

* No segmentation provided by the radiologist.

Figure 4.8 shows examples of model's false detections, where areas in the image are detected as tumor, though they are healthy tissues (probably lymph nodes). These areas have image contrast quite similar to the breast tumors, which leads to their false detection. Moreover, there are other healthy tissues in the breast area that may resemble the tumor in DCE images. Figure 4.9 shows examples in which the model has detected nipple and breast veins incorrectly.

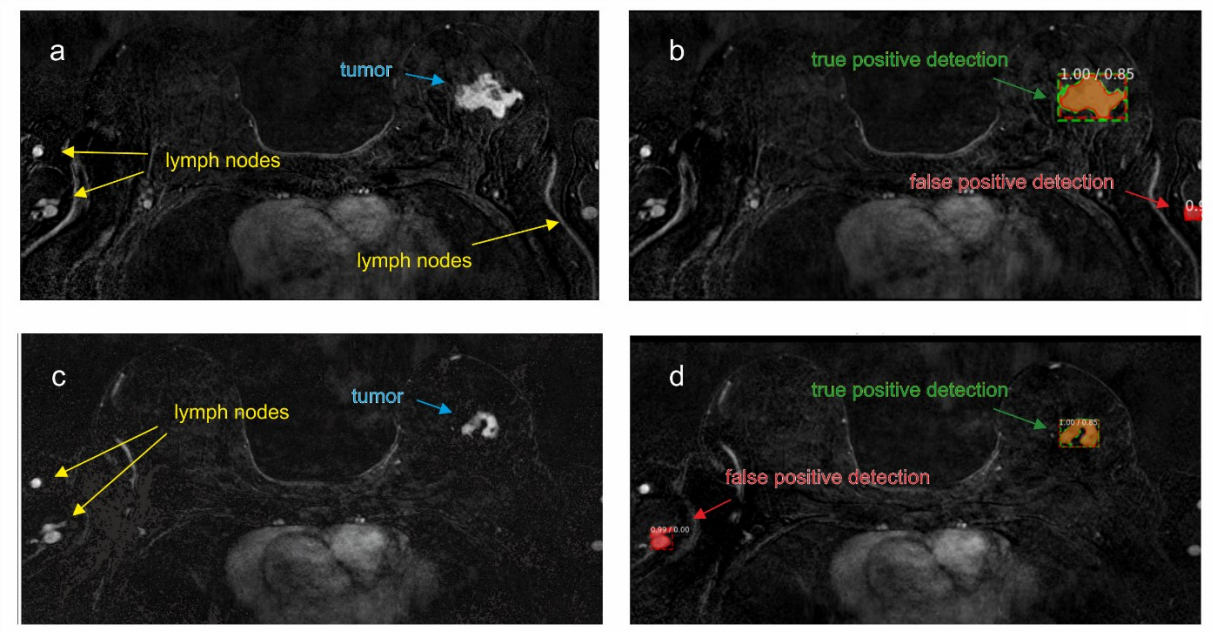


Figure 4.8: Examples of false positive detections of the lymph nodes as breast cancer tumors in Trondheim (test) dataset. Images on the left are original DCE-MR images and images on the right are model's outputs that contain false positive detections of the lymph nodes.

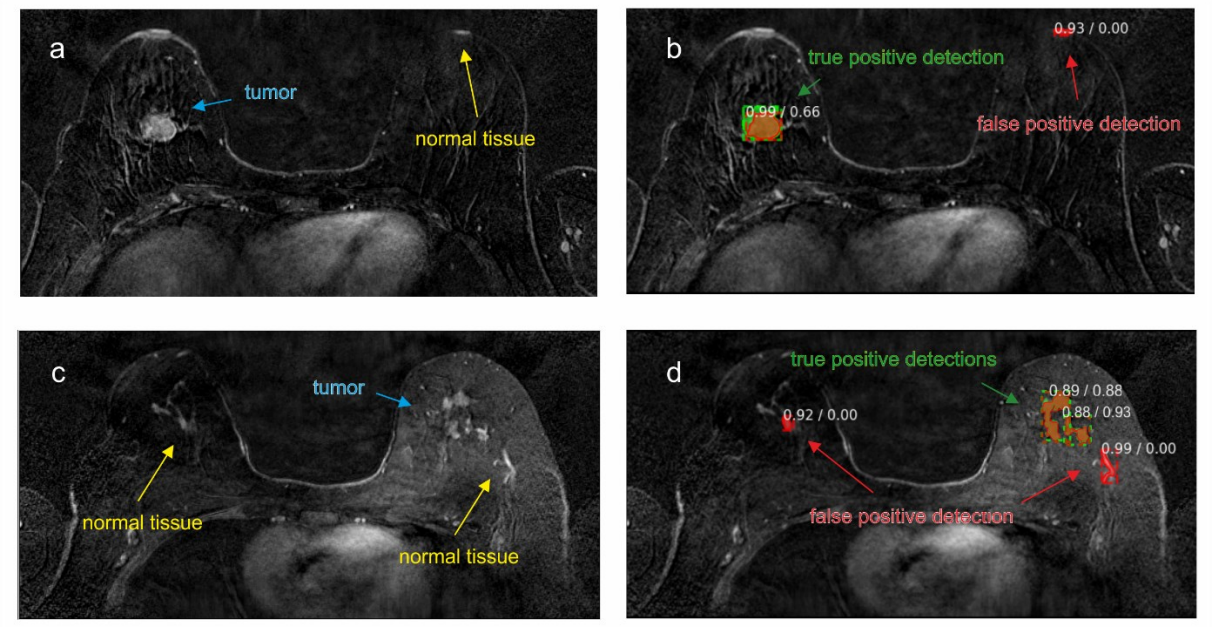


Figure 4.9: False positive detections of normal tissues by the model which can be due to training the model by images that contain small individual masses of tumor. (a) patient with a tumor in the right breast. (b) true positive detection of the tumor and false positive detection of the nipple as a tumor. (c) patient with multiple tumor masses in her left breast. (d) two true positive detections and two false positive detections of the normal breast tissues that are likely veins.

Overall, the model did not detect tumors in 23 slices of the test set. Most of these slices (17/23) were first/last slices of the tumor volume that contain very small area of the tumor (see Figure 4.2) and the

rest were images that lack sufficient quality. These detections are considered as false negatives that decrease sensitivity and accuracy of the model, since these values are a binary classification and are not weighted by tumor size within the slice. Figure 4.10 shows examples of the model's false negative detections.

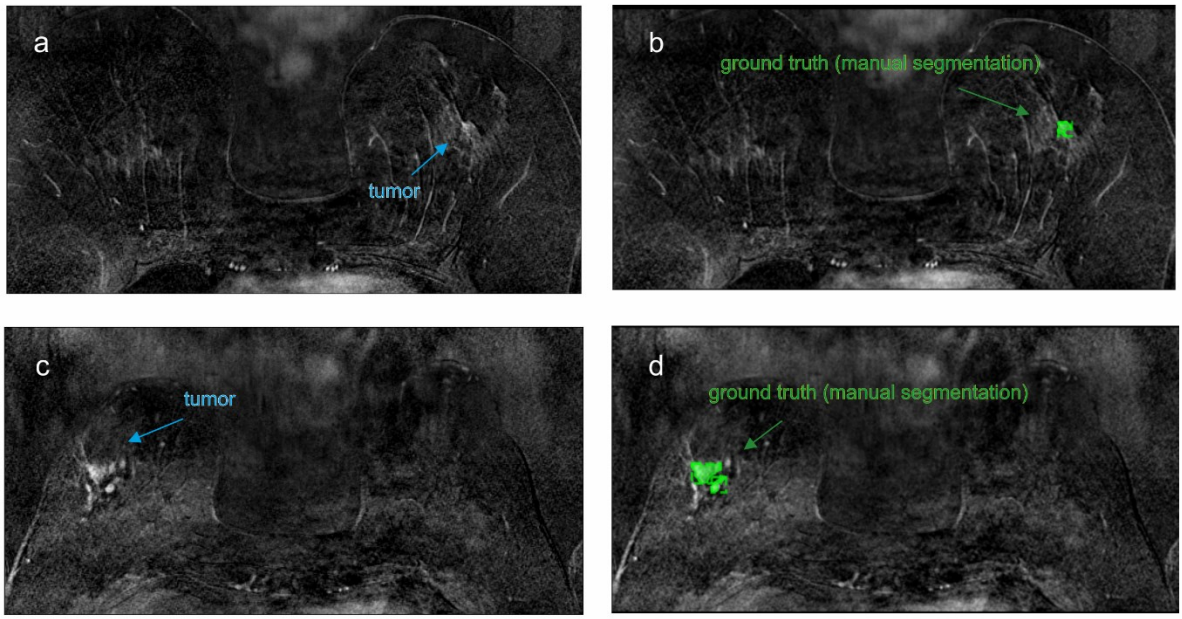


Figure 4.10: False negative detections by the model. (a) First slice of the tumor volume containing very small area of the tumor in the left breast of the patient. (b)A false negative detection by the model, where only the ground truth is visualized by the model. (c) A DCE image with low CNR, showing tumor in the right breast of the patient. (d) Inability of model in detection of the tumor, which is considered as a false negative detection.

Chapter 5

Discussion

In this work, the Mask R-CNN machine learning algorithm was applied to DCE MRI data taken from a clinical breast cancer trial. The model was trained using researcher-drawn tumor segmentations, which were found to be an acceptable proxy for radiologist-validated 'gold standard' segmentations. Comparison of the model results to the manual segmentation was generally in good concordance, although there were a number of false results, both positive and negative. The results show that deep learning models can be successfully trained to detect and segment tumors. Overall, the method showed good potential for automated segmentation of breast cancer tumors, which may have significant impact on radiologist workload.

Machine learning is an exciting topic in radiology, but it is associated with a number of challenges. Data preparation is one of the most demanding steps in any machine learning project. In this work, approximately 5000 images were segmented manually for training and evaluation of the Mask R-CNN. Performing this process for such a large quantity of data is highly time-consuming, tedious, and subjective. To define each image in the dataset and its corresponding ground truth for the Mask R-CNN model, a meta data file that contains this information should be fed as an input to the model. In case of ordinary images (such as JPEG, PNG, etc.), there are many different image annotator softwares that output this meta data file, which can be used directly in Mask R-CNN. Unfortunately, these softwares do not accept medical image formats such as NIFTI or DICOM, and ITK-SNAP is not capable of creating such a file. Therefore, an in-house MATLAB script was used to convert the generated masks by ITK-SNAP into an acceptable format by the model.

The other option was converting the medical images from DICOM format into a format admissible by available softwares. But converting DICOM images to formats like JPEG or PNG leads to loss of image quality due to compression artifacts or makes them unusable by decreasing image depth to 8-bit, which is insufficient for medical images. Instead, the images should be converted to a standard research format such as NIFTI or directly to numpy arrays in Python to preserve the quality of the images. Nifti format stores essential 3D meta data of the medical images, thus it can be used as an alternative to DICOM images when we only need the image data.

In order to train a model optimally, different confidence thresholds and hyper-parameters have to be tuned. Since tumor voxels in DCE images in this work are given as binary classification, detections with normalized predicted probabilities less than the user-defined threshold are assigned to class 0

(background) and values greater than the threshold are assigned to class 1 (tumor). Thus, different thresholds lead to different TP, FP, TN and FNs and consequently different reported accuracies of the model.

In this work, the model confidence threshold was set to 0.85, meaning that an object in the image will be assigned to the tumor class if the probability is greater than 0.85, otherwise it will be classified as background. Low thresholds let the model classify more objects as tumor in the image, which can increase sensitivity on a voxel level but also leads to an increase in false positive detections, eventually decreasing the model's accuracy. In contrast, increasing the model's threshold eliminates low probability detections, which reduces both true and false positives and increases true and false negatives. Additionally, there are many hyper-parameters in a deep learning model and finding the best configuration to optimize the model in such a high-dimensional space is not trivial. Setting the number of hidden layers, learning rate, learning momentum, and weight decays as some of the hyper-parameters, requires extensive trial and error for achieving the best performance of the model. A full optimization of the model was not feasible for this work, due to the short time of having access to GPU in HUNT-cloud, although a limited test of the threshold value suggested 0.85 as a suitable value.

The PeTreMac dataset is comprised of three independent datasets from collaborating hospitals in Bergen, Stavanger, and Trondheim. These datasets have different imaging parameters such as resolution, slice thickness, and magnetic strength of the MR machine which give differences between their images. For example, DCE-MR images of the Bergen and Stavanger have isotropic in-plane resolution with same width and height, whereas the Trondheim dataset has wider images with 2:1 aspect ratio (table 3.1 and Figure 3.2). In addition, contrast to noise ratio of Trondheim dataset is lower than the other datasets, which makes it more difficult for the model to detect and segment the tumors in test set. Where possible, the training data should be a reflection of the expected test data, and this can be achieved by using larger datasets or careful data collection and curation, with both approaches having practical challenges.

In several images of the Trondheim dataset there are visible lymph nodes that have similar intensity to the breast tumors but in locations close to the lateral edges of the image (Figure 4.8). In some cases, these are detected as tumor by the model, leading to an increase in false positive rates. Cropping the images to the breast area and/or adding part of the Trondheim dataset to the training set could decrease the false positive of lymph nodes. Conversely, the association of lymph nodes with invasiveness of cancer is a potentially important target for automated detection through machine learning, and further work could investigate the performance as applied to lymph nodes.

The other factor that increases the rate of false positive detections arises from training the model by images that contain widely spread-out tumors. Such tumors contain multiple individual small masses that need to be segmented separately, which resemble small white areas in the DCE images of normal breasts that might be detected by the model (Figure 3.2c). An example of such a false positive detection can be seen in Figure 4.9d; and suggests that the model may perform less well for certain types of tumors.

In case of false negative detections, the model did not classify any voxels as the tumor in the image. The majority (74%) of these false negatives were related to the first/last slices of the tumor that contain very small areas of tumor volume (Figure 4.2), and the remaining false negatives were images with very poor CNRs (Figure 4.10). False negatives are very important in medical screening, because diagnosing a sick person as healthy may lead to severe complications in future but diagnosing a healthy person as sick can

be reclaimed by additional tests. It is reasonable to conclude that minimum CNR is required for automated segmentation, and that images could be tested before use with the model.

Regarding dice similarity coefficients between researcher-drawn and radiologist-validated segmentations in the Trondheim dataset, images with high CNR and slices with round-shaped tumors had the highest DSCs (Figure 4.3). Conversely, lowest average DSCs belong to the slices with low CNRs, tumors with irregular shapes and tumor-containing slices with multiple small masses of the tumor (Figure 4.4). Comparing DSC values between Table 4.1 and Table 4.2 shows that DSCs of the model segmentations is in total accordance with the DSCs of researcher-drawn/radiologist-validated segmentations; average DSCs between researcher-drawn and radiologist-validated segmentations have their lowest values for the same patients as for the DSCs between model and researcher-drawn/radiologist-validated segmentations. This applies to highest average DSCs too.

There is also an outlier value in distribution of average DSC between model and manual segmentations (Figure 4.6), with the lowest average DSC (0.57) among all the patients and indicates a very poor performance in this case. This tumor was not segmented by the radiologist, indicating that this poor performance is related to the difficulty of the case.

While there are many factors to consider in the implementation of machine learning as a technology, the use must also be considered in the context of added value to the clinical process and patient management. This will ultimately require smooth integration into the radiologist's workflow, and assessment as to whether the technology can truly be used unsupervised, or whether it is used as a tool to streamline, but not ultimately replace, the radiologist role. Challenges for adoption into everyday clinical use include integration into the normal image-reviewing softwares (e.g. PACS system), as well as well-documented validation studies and appropriate setting that might be biases towards sensitivity or specificity (depending on the specific application). The work in this thesis shows a reasonable success from a relatively small, local dataset (a single clinical trial), but more accumulation and sharing of data in the future will allow training datasets that will cover most variations of demographics, scanner hardware and scanning protocols, leading to increased performance. It should also be considered that while the radiologist segmentation is considered the 'gold standard', this is not necessarily the underlying truth, and the true target of machine learning is not just to replicate the radiologist segmentation, but to improve on it.

Chapter 6

Conclusion and future work

6.1 Conclusion

Dynamic contrast-enhanced MRI has shown to have the highest sensitivity in detection of the residual breast tumor following neoadjuvant chemotherapy. To evaluate the tumor response to the treatment, radiologists segment the tumors in MRI slices manually to measure the volume changes after each treatment cycle. This task is highly time consuming and often has a large variability leading to different estimations of the tumor volume.

The goal of the thesis was to implement a deep learning model for automatic detection and segmentation of tumors in locally-advanced breast cancers using DCE-MR images to evaluate the treatment response post neoadjuvant chemotherapy. For this purpose, Mask R-CNN, a region-based CNN that performs object detection and instance segmentation was used.

The results of the model in detection of the breast tumors, trained and tested on a reasonable small and heterogeneous dataset, show that Mask R-CNN is capable of detecting breast tumors in DCE-MR images with a high accuracy (0.84). In addition, the dice similarity coefficient between radiologist-validated segmentations and the model predictions was 0.83, which is promising given that the training segmentations were not radiologist-validated, and suggests that the quality of the segmentations could be further improved. The use of the machine learning tool requires an investment of time in providing the 'ground truth' values from which to learn, but once trained can potentially reduce the workload for radiologists.

6.2 Future work

To improve the model's performance in detection of the breast tumors, adding more training data with different CNRs and aspect ratios is likely beneficial. It can be investigated by adding part of the Trondheim dataset to the training set, to let the model learn from images with different resolution and CNRs. Moreover, the performance of the model can be increased by tuning the hyper-parameters such as learning rate or number of hidden layers. Another method to improve the model's performance is to perform pre-processing on the input data like normalization, which brings the variations to a specific range.

Mask R-CNN can also classify different objects in an image. In this thesis, since the patient cohort was locally-advanced breast cancers, detected objects were classified in only one class, which was tumor. In a future work, by adding more data, Mask R-CNN can be applied to classify tumors into benign and malignant ones, based on their contrast, size and shape. It would also be interesting to investigate if the Mask R-CNN could be used to subtype the malignant tumors according to their molecular profiles, although for these more complex classification problems the amount of training is proportionately larger

Since the DCE-MRI has the highest sensitivity in breast cancer detection, and is a core component of breast cancer screening MR, it would be interesting to test the model in a screening cohort. Since the model in this work was trained by using MR images of women with locally-advanced breast cancer, it would need further tuning and adding images of screening cohort into the training process for this purpose.

Bibliography

[1] Cancer Registry of Norway (2019) Cancer in Norway 2019— cancer incidence, mortality, survival and prevalence in Norway. Accessed 17 March 2021.

[2] Price, E. R., Wong, J., Mukhtar, R., Hylton, N., & Esserman, L. J. (2015). How to use magnetic resonance imaging following neoadjuvant chemotherapy in locally advanced breast cancer. *World journal of clinical cases*, *3*(7), 607–613. https://doi.org/10.12998/wjcc.v3.i7.607.

[3] Kwasigroch, Arkadiusz & Mikołajczyk, Agnieszka & Grochowski, Michał. (2017). Deep convolutional neural networks as a decision support tool in medical problems – malignant melanoma case study. 848-856. 10.1007/978-3-319-60699-6_81.

[4] Wang, Dayong & Khosla, Aditya & Gargeya, Rishab & Irshad, Humayun & Beck, Andrew. (2016) Deep Learning for Identifying Metastatic Breast Cancer, Quantitative Methods; Computer Vision and Pattern Recognition, arXiv:1606.05718.

[5] Kwasigroch, Arkadiusz & Jarzembinski, Bartlomiej & Grochowski, Michał. (2018). Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy. 111-116. 10.1109 /IIPHDW.2018.8388337.

[6] Breast cancer: Clinical case, prophylaxis and diagnosis. Available from: https://www.kenhub.com/ en/library/anatomy/breast-cancer-development-after-prophylactic-subcutaneous-mastectomy. Accessed 12 Feb 2021.

[7] After a Biopsy: Making the Diagnosis. Available from: https://www.cancer.net/navigating-cancer-care/ diagnosing-cancer/reports-and-results/after-biopsy-making-diagnosis. Accessed 16 Feb 2021.

[8] LCIS, ADH and ALH may increase your risk, Breast Screen Victoria, Lobular carcinoma in situ, atypical lobular hyperplasia and atypical ductal hyperplasia of the breast. Available from: https://www.breastscreen.org.au/breast-cancer-and-screening/your-breast-cancer-risk/lcis-adh-and-alh/. Accessed July 2018.

[9] Middleton, L. P., Sneige, N., Coyne, R., Shen, Y., Dong, W., Dempsey, P., & Bevers, T. B. (2014). Most lobular carcinoma in situ and atypical lobular hyperplasia diagnosed on core needle biopsy can be managed clinically with radiologic follow-up in a multidisciplinary setting. Cancer medicine, 3(3), 492–499. https://doi.org/10.1002/cam4.223.

[10] Most Common Molecular Subtypes of Breast Cancer. Available from: https://www.cancercenter. com/cancer-types/breast-cancer/types/breast-cancer-molecular-types. Accessed 16 Feb 2021.

[11] Stages of breast cancer - Canadian Cancer Society. Available from: https://www.cancer.ca/en/ cancer-information/cancer-type/breast/staging/?region=qc. Accessed 16 Feb 2021.

[12] Breast Cancer Staging - Jordan Breast Cancer Program. Available from: http://www.jbcp.jo/ understandingbreastcancer/33. Accessed 16 Feb 2021.

[13] Thompson AM, Moulder-Thompson SL. Neoadjuvant treatment of breast cancer. Ann Oncol. 2012 Sep;23 Suppl 10(Suppl 10):x231-6. doi: 10.1093/annonc/mds324. PMID: 22987968; PMCID: PMC6278992.

[14] Moo TA, Sanford R, Dang C, Morrow M. Overview of Breast Cancer Therapy. PET Clin. 2018 Jul;13(3):339-354. doi: 10.1016/j.cpet.2018.02.006. PMID: 30100074; PMCID: PMC6092031.

[15] Knopp MV, Weiss E, Sinn HP, Mattern J, Junkermann H, Radeleff J, Magener A, Brix G, Delorme S, Zuna I, van Kaick G. Pathophysiologic basis of contrast enhancement in breast tumors. J Magn Reson Imaging. 1999 Sep;10(3):260-6. doi: 10.1002/(sici)1522-2586(199909)10:3<260::aid-jmri6>3.0.co;2-7. PMID: 10508285.

[16] Nnewihe, A. N., Hargreaves, B. A., Daniel, B., Gold, G. E., & Stanford University. (2012). High resolution breast MRI.

[17] American College of Radiology Practice Guidelines for the Performance of Magnetic Resonance Imaging of the Breast. Available at: http:// www.acr.org. Accessed on September 13, 2010.

[18] Uwe Fischer and Ulrich Brinch(2004). Practical MR mammography. Georg Thieme Verlag.

[19] Daniel Förnvik (2008). Complementary analysis of breast cancer using MRI and breast tomosynthesis.

[20] Loiselle, C., Eby, P. R., Kim, J. N., Calhoun, K. E., Allison, K. H., Gadi, V. K., Peacock, S., Storer, B. E., Mankoff, D. A., Partridge, S. C., & Lehman, C. D. (2014). Preoperative MRI improves prediction of extensive occult axillary lymph node metastases in breast cancer patients with a positive sentinel lymph node biopsy. Academic radiology, 21(1), 92–98. https://doi.org/10.1016/j.acra.2013.10.001

[21] Lord SJ, Lei W, Craft P, Cawson JN, Morris I, Walleser S, Griffiths A, Parker S, Houssami N. A systematic review of the effectiveness of magnetic resonance imaging (MRI) as an addition to mammography and ultrasound in screening young women at high risk of breast cancer. Eur J Cancer. 2007 Sep;43(13):1905-17. doi: 10.1016/j.ejca.2007.06.007. Epub 2007 Aug 2. PMID: 17681781.

[22] Obdeijn IM, Winter-Warnars GA, Mann RM, Hooning MJ, Hunink MG, Tilanus-Linthorst MM. Should we screen BRCA1 mutation carriers only with MRI? A multicenter study. Breast Cancer Res Treat. 2014 Apr;144(3):577-82. doi: 10.1007/s10549-014-2888-8. Epub 2014 Feb 25. PMID: 24567197.

[23] Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin. 2007;57:75-89.

[24] M. T. Vlaardingerbroek and J. A. den Boer (2010). Magnetic Resonance Imaging: Theory and Practice. Springer Verlag, 3rd edition.

[25] Müller, D., Kramer, F. MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. BMC Med Imaging 21, 12 (2021). https://doi.org/10.1186/s12880-020-00543-7.

[26] Patel, Rachana & Patel, Sanskruti. (2020). A Comprehensive Study of Applying Convolutional Neural Network for Computer Vision. International Journal of Advanced Science and Technology. 6. 2161-2174.

[27] Albelwi, S.; Mahmood, A. (2017). A Framework for Designing the Architectures of Deep Convolutional Neural Networks. *Entropy*, *19*, 242. https://doi.org/10.3390/e19060242.

[28] Waleed Abdulla. Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN, 2017.

[29] He K, Gkioxari G, Dollár P, Girshick R. (2017). Mask R-CNN. IEEE Trans Pattern Anal Mach Intell. 2020;42(2):386–97.

[30] Ren S, He K, Girshick R, Sun J. Faster R-CNN (2016) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, IEEE Trans Pattern Anal Mach Intell. 2017;39(6):1137–49.

[31] Vishnu Subramanian, Deep Learning with PyTorch: A practical approach to building neural network models using PyTorch, Packt Publishing, 2018, ISBN-10: 1788624335.

[32] Teruel, J. R., Heldahl, M. G., Goa, P. E., Pickles, M., Lundgren, S., Bathen, T. F., and Gibbs, P. (2014), Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer, NMR Biomed., 27, pages 887– 896, doi: 10.1002/nbm.3132.

[33] Adrian Rosebrock, Deep Learning for Computer Vision with Python (2017). PYIMAGESEARCH, 1st edition.

[34] Ian Goodfellow, Yoshua Bengio, Aaron Courville (2016). Deep Learning, Adaptive Computation and Machine Learning series, The MIT Press.

[35] MIT License. URL: https://opensource.org/licenses/MIT.

[36] M. T. Vlaardingerbroek and J. A. den Boer (2010). Magnetic Resonance Imaging: Theory and Practice. Springer Verlag, 3rd edition.

Appendix A: model configuration

Configurations:

BACKBONE	resnet101
BACKBONE_STRIDES	[4, 8, 16, 32, 64]
BATCH_SIZE	1
BBOX_STD_DEV	[0.1 0.1 0.2 0.2]
COMPUTE_BACKBONE_SHAPE	None
DETECTION_MAX_INSTANCES	100
DETECTION_MIN_CONFIDENCE	0.85
DETECTION_NMS_THRESHOLD	0.3
FPN_CLASSIF_FC_LAYERS_SIZE	1024
GPU_COUNT	1
GRADIENT_CLIP_NORM	5.0
IMAGES_PER_GPU	1
IMAGE_CHANNEL_COUNT	3
IMAGE_MAX_DIM	1024
IMAGE_META_SIZE	14
IMAGE_MIN_DIM	800
IMAGE_MIN_SCALE	0
IMAGE_RESIZE_MODE	square
IMAGE_SHAPE	[1024 1024 3]
LEARNING_MOMENTUM	0.9
LEARNING_RATE	0.001
LOSS_WEIGHTS	{'rpn_class_loss': 1.0, 'rpn_bbox_loss': 1.0, 'mrcnn_class_loss': 1.0, 'mrcnn_bbox_loss': 1.0, 'mrcnn_mask_loss': 1.0}
MASK_POOL_SIZE	14
MASK_SHAPE	[28, 28]
MAX_GT_INSTANCES	100
MEAN_PIXEL	[123.7 116.8 103.9]
MINI_MASK_SHAPE	(56, 56)
NAME	tumor_detector
NUM_CLASSES	2
POOL_SIZE	7
POST_NMS_ROIS_INFERENCE	1000
POST_NMS_ROIS_TRAINING	2000
PRE_NMS_LIMIT	6000
ROI_POSITIVE_RATIO	0.33
RPN_ANCHOR_RATIOS	[0.5, 1, 2]
RPN_ANCHOR_SCALES	(32, 64, 128, 256, 512)
RPN_ANCHOR_STRIDE	1
RPN_BBOX_STD_DEV	[0.1 0.1 0.2 0.2]

RPN_NMS_THRESHOLD	0.7
RPN_TRAIN_ANCHORS_PER_IMAGE	128
STEPS_PER_EPOCH	50
TOP_DOWN_PYRAMID_SIZE	256
TRAIN_BN	False
TRAIN_ROIS_PER_IMAGE	200
USE_MINI_MASK	True
USE_RPN_ROIS	True
VALIDATION_STEPS	50
WEIGHT_DECAY	0.0001

NTNU

Norwegian University of
Science and Technology