Die approbierte Originalversion dieser Diplom-/Masterarbeit ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.
http://www.ub.tuwien.ac.at

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology.
http://www.ub.tuwien.ac.at/eng

FAKULTÄT
FÜR !NFORMATIK

Faculty of Informatics

Anomaly Detection and Prediction in longitudinal Imaging Data

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Biomedical Engineering

eingereicht von

Bianca Burger,
Matrikelnummer 01225150

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Robert Sablatning
Mitwirkung: Assoc.Prof. Dipl.-Ing. Dr.techn. Georg Langs

Wien, 23. Juli 2018

Bianca Burger	Robert Sablatning

FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Anomaly Detection and Prediction in longitudinal Imaging Data

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Biomedical Engineering

by

Bianca Burger,
Registration Number 01225150

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Robert Sablatning
Assistance: Assoc.Prof. Dipl.-Ing. Dr.techn. Georg Langs

Vienna, 23rd July, 2018

Bianca Burger

Robert Sablatning

Erklärung zur Verfassung der Arbeit

Bianca Burger,
Ollersbachsiedlung 32, 2261 Angern

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 23. Juli 2018

Bianca Burger

Danksagung

An dieser Stelle möchte ich meine Dankbarkeit allen Personen aussprechen, die mich bei dieser Arbeit unterstützt haben.
Als aller erstes möchte ich Georg Langs vom Computational Imaging Research Lab (CIR) an der Medizinischen Universität Wien danken für die Betreuung meiner Arbeit, vom Festlegen des Themas bis zum Korrekturlesen am Ende. Außerdem möchte ich ihm für das Wecken meines Interesses an medizinischer Bildverabeit im gleichnamigen Kurs an der TU Wien danken under der Möglichkeit, dass ich meine Diplomarbeit am CIR schreiben durfte.
Ich möchte mich auch bei Robert Sablatnig vom Computer Vision Lab an der Technischen Universität Wien bedanken für seine konstruktive Kritik an meiner Arbeit. Außerdem möchte ich mich bei ihm und Amy Bruno-Lindner für das im Kurs „Scientific Presentation and Communication“ vermittele Wissen bedanken.
Weiters möchte ich mich bei Markus Holzer, Wolf-Dieter Vogl und Thomas Schlegl für die Hilfe bei der Datenübertragung, bei der Konvertierung der Daten von DICOM- zu Nifti-Format und der Beantwortung von Fragen zu Generative Adversarial Networks bedanken.
Bei Thomas Helbich, Katja Pinker und Maria Bernathova bedanke ich mich für die MRT-Daten, die mir zur Verfügung gestellt wurden und die klinische Kooperation. Ich bedanke mich außerdem bei Verena Stanzl und Thomas Deimel für die gute Arbeitsatmosphäre im Büro.
Zu aller letzt möchte ich mich speziell bei meinen Eltern bedanken, die mich nicht nur während des Schreibens der Abschlussarbeit und dem Studium unterstützt haben, sondern mein ganzes Leben. Ohne euch würde ich nicht da stehen, wo ich heute stehe. Vielen Dank dafür!

Acknowledgements

At this point I would like to express my gratitude to all persons who supported me during the work on this thesis.

First of all I would like to thank Georg Langs from the Computational Imaging Research Lab (CIR) at the Medical University of Vienna for his guidance through the whole thesis, from determining the aim of the work to the proof-reading at end. I also want to thank him for raising my interest in medical image processing in the course of the same name at Vienna University of Technology and giving me the opportunity to write my thesis at the CIR.

I also want to thank my supervisor Robert Sablatning from the Computer Vision Lab at Vienna University of Technology for his constructive comments on my work. Furthermore, I want to thank him and Amy Bruno-Linder for the knowledge conveyed in the course "Scientific Presentation and Communication".

Further, I want to thank Markus Holzer, Wolf-Dieter Vogl and Thomas Schlegl for supporting me in the data transfer, the conversion from DICOM- to Nifti format and getting started with Generative Adversarial Networks. I also thank Thomas Helbich, Katja Pinker and Maria Bernathova for the provided MRI data and the clinical cooperation. I want to thank Verena Stanzl and Thomas Deimel for the good working atmosphere in the office.

My final and special thanks go to my parents, who not only supported me during the thesis and my studies, but through my entire life. Without you, I didn't get that far, thank you!

Kurzfassung

Generative Adversarial Network (GANs) fanden schon in verschiedenen Bereichen Anwendung wie Feature Extraktion, Bildbearbeitung und das Detektieren von krankhaften Gewebe der Retina. In dieser Arbeit wird ein GAN verwendet, um einen Anomalie-Score zu erstellen und damit Läsionen in Hochrisikopatientinnen für Brustkrebs zu erkennen. Der Anomalie-Score quantifiziert dabei die Abweichung von der Verteilung gesunden Gewebes, wobei Bilder, die in einem Zeitraum von 2 bis 13 Jahren durch Perfusions-Magnetresonanztomographie erstellt wurden, für das Training des GAN herangezogen werden.

Die Methodik besteht im Wesentlichen aus 3 Stufen, nämlich der Intra-Subjekt Registrierung, der Segmentierung des Brustgewebes und schließlich aus der eigentlichen Anomaliedetektion. Die Registrierung ist notwendig um innerhalb einer Patientin örtliche Korrespondenz der Bilder zu erhalten und die Segmentierung wird benötigt um die Positionen der Voxel, die Brustgewebe entsprechen, in Erfahrung zu bringen. Der Anomalie-Score wird dann basierend auf Differenzenbildern zwischen zwei Aufnahmezeitpunkten einer Patientin und dessen Vergleich mit dem GAN-Model von gesundem Gewebe berechnet.

Die Evaluierung der in dieser Arbeit vorgestellten Herangehensweise an Bildausschnitten von 5 gesunden und 8 kranken Patientinnen zeigt eine Sensitivität von 99.5% und eine Spezifität von 84%. Das Vorhandensein von Läsionen wurde zuvor von Radiologen und Radiologinnen diagnostiziert. Außerdem zeigen die Resultate auch, dass es mit Hilfe des Anomalie-Scores möglich ist Läsionen einen Aufnahmezeitpunkt früher zu erkennen als der Radiologe oder die Radiologin. In diesem Fall liegt die Sensitivität bei 92.7% und die Spezifität bei 78.6%.

Abstract

Generative Adversarial Networks (GANs) have been applied in different fields including feature extraction, image inpainting and the detection of abnormal tissue of the retina. In this thesis a GAN is used to create an anomaly score to detect lesions in high risk breast cancer patients, whereas the score represents the divergence from healthy tissue. Dynamic Contrast-Enhanced Magnetic Resonance (DCE-MR) images for each patient, acquired in a time period ranging from 2-13 years serve as information source for GAN training.

The methodology consists of 3 main steps, namely intra-patient image registration, whole breast segmentation and the anomaly detection itself. Images belonging to the same subject are brought into the same coordinate frame to reach spatial correspondence across time. Breast segmentation is necessary to obtain the positions of breast tissue voxels. The anomaly score is then calculated based on difference images between the acquired time points of a patient, by a comparison with the GAN model of normal breast tissue. Evaluation of the detection performance on image patches from 5 healthy and 8 diseased patients demonstrates a sensitivity of 99.5% and a specificity of 84%. The presence of lesions has been confirmed by a radiologist. Furthermore, results show that the score allows the identification of lesions acquired one time point earlier than they have been identified by a radiologist. The sensitivity in this case is 92.7% and the specificity is 78.6%

Contents

Kurzfassung **xi**

Abstract **xiii**

Contents **xv**

1 Introduction **1**
1.1 Problem Statement 1
1.2 Aim of the work 2
1.3 Methodological Approach 2
1.4 Thesis outline 3

2 Breast MR Imaging **5**
2.1 Basics of Nuclear Spin 5
2.2 Basics of Tomography 6
2.3 Breast MRI 8
2.4 Discussion 10

3 Image Registration **11**
3.1 Constituents of a Registration Framework 11
3.2 Examples of Matching Criteria: Mutual Information and Cross-Correlation 12
3.3 Examples of Deformation Models: SyN and Demons 13
3.4 Optimization Strategy 17
3.5 Discussion 17

4 Generative Adversarial Networks **19**
4.1 The idea behind Generative Adversarial Networks 19
4.2 Deep Networks 22
4.3 Deep Convolutional Generative Adversarial Nets 32
4.4 Wasserstein GAN and further improvements for GAN-Trainig 34
4.5 Discussion 36

5 Specific Related Work **37**
5.1 Breast Registration 37

5.2 Breast Segmentation 38
5.3 Anomaly and Lesion Detection 38
5.4 Discussion 39

6 Methodology **41**
6.1 Data and Preprocessing 41
6.2 Image Registration 43
6.3 Whole Breast Segmentation 46
6.4 Anomaly detection 48
6.5 Discussion 53

7 Experiments and Results **55**
7.1 Initial Experiments to choose Registration Algorithm 55
7.2 Evaluation of registration 60
7.3 Evaluation of Segmentation 62
7.4 Evaluation of Anomaly Detection 65
7.5 Evaluation of Anomaly Prediction 76

8 Conclusion **79**
8.1 Summary 79
8.2 Future Work 80

Bibliography **83**

CHAPTER 1

Introduction

Medical imaging is a tool for detecting and diagnosing breast cancer [22]. In this context, for high risk patients regular control of their health status is necessary and MRI is an imaging technique for detecting lesions in the breast [22], [10]. There are different MRI techniques, with each of them making use of a certain property of tumour cells, so that lesions differ from healthy tissue in images [22]. Dynamic Contrast-enhanced MRI (DCE-MRI) alone has an excellent sensitivity with negative-predictive values ranging between 89% and 99% and a good specificity ranging from 47% to 97% [22]. Therefore it serves as backbone for any breast MRI. To increase the specificity, DCE-MRI is combined with other MRI techniques to a multi-parametric approach [22]. Such an additional parameter is for example diffusion weighted imaging. However, this thesis focuses on DCE-MR imaging data, since they are available with consistent image acquisition parameters over time.

In practice, images need to be examined for anomalies [10]. This can be done by radiologists but manually investigation depends on the expertise of the interpreter[10]. Algorithms that can detect anomalies automatically save time used for manually interpretation of 4D DCE-MRI data and provide the radiologist with information about possible lesion locations [10].

1.1 Problem Statement

For high risk breast cancer patients DCE-MRI is used for regular screening, leading to longitudinal breast imaging data of these patients [22]. This allows using longitudinal data for lesion detectors. Furthermore, this temporally linked data can be used to evaluate how early a model is able to detect lesions compared to the radiologist who annotated the imaging data. Anomaly detectors which notice the presence of a suspicious area before a radiologist can, are highly desirable.
The first goal of this thesis is to create a model that can detect anomalies in imaging

data as well as the examination of its abilities to identify lesions in the breast, based on follow-up sequences. The second goal is to evaluate, this model's capability to predict the emergence of lesions.

1.2 Aim of the work

The goal of this thesis is the detection of anomalies in MRI images of the breast based on temporally linked imaging data. Breasts of healthy women vary in their shape, depending on age, diet and choice of clothes [24]. The combined information at different time points of healthy women are therefore used to train a Generative Adversarial Network to learn the variance in appearance of healthy breast tissue. This knowledge is then used to find suspicious regions in high-risk breast cancer patients by identifying their images as deviating from normal appearance.
Temporally linked data for each woman is used.

The main contributions are:

- training a GAN on MRI data of the breast using temporally linked images
- using follow up data of high risk breast cancer patients collected over several years as information source for anomaly detection
- adapting a GAN- based anomaly score for the detection of breast lesions
- adapting a GAN-based anomaly score to detect lesions before a radiologist has detected them

1.3 Methodological Approach

The methodical approach consists of the following steps:

- **Intra-patient registration**
  For each person, all images available across time points are registered to the same image frame to reach spatial correspondence. Images with different parameters but acquired at the same time are already co-registered. Therefore the focus is on inter-time point registration.
- **Segmentation of the breast**
  After registration, one breast mask per patient is calculated to obtain the regions of breast tissue. Only image patches showing parts of the breast are used for further analysis.

- **Anomaly detection**
    - Model Training: Patches are extracted from difference images calculated between time points and serve as information source for GAN training. Only healthy patients contribute to the training set. After completed training the generating part of the GAN is able to reproduce samples of healthy patches. This is used to detect anomalies.
    - Anomaly Detection: New images are down-sampled to the input space of the generating part of the GAN and a rebuilt version is created by the Generator. This artificial patch is compared with the original one in two ways. The first one is the calculation of image residuals. The second one is calculating feature residuals by comparing the features of the real and the generated patch. Those two pieces of information are combined in a weighted anomaly score. Based on this score, artificial images which diverge to much from the original image are classified as containing a lesion.
- **Prediction**
  The procedure to predict lesions is the same as for the detection, however the test sets are composed differently.
- **Evaluation**
  The ability of the GAN-based anomaly score to detect and predict anomalies is evaluated on a test set consisting of an equal amount of images from healthy women and breast cancer. For detection, difference images which have been calculated on time points with confirmed lesions are used. To evaluate the prediction performance, instead of the time point with a lesion, one time point earlier is used.

1.4 Thesis outline

The thesis consists of 8 chapters and is structured as follows:

Chapter 1 is the introduction and contains the motivation for and the aim of this thesis. Besides, it includes an overview of the methodological approach used in this thesis as well as the thesis outline.

Chapter 2: Breast MR Imaging deals with the physical basics of magnetic resonance imaging and its usage in breast imaging.

Chapter 3: Image Registration gives a theoretical overview of a registration framework. For each part of the framework examples, which are used or mentioned in this thesis, are given.

Chapter 4: Generative Adversarial Networks discusses the idea behind *Generative Adversarial Networks* (GANs) and its building blocks *Neural Networks.* Furthermore, the chapter goes through the papers which have led to the use of GANs in anomaly detection and provides the reader with possibilities to improve GAN-training.

Chapter 5: Specific Related Work lists related work for the three corner stones of this thesis, namely intra-patient registration, whole breast segmentation and anomaly detection.

Chapter 6: Methodology describes the methodological approach proposed in this thesis. It includes some words concerning the preprocessing of the patients' data and leads through the approaches used for intra-patient registration, whole breast segmentation and lesion detection. The latter is subdivided into GAN-training and setting up an anomaly score.

Chapter 7: Experiments and Results is divided into 4 main parts. The first two sections describe experiments which have led to the registration pipeline as described in Chapter 6 and the evaluation of the pipeline. The next part describes some experiments concerning whole breast segmentation. The final part deals with the evaluation of the anomaly detection and prediction performance using the anomaly score which has been set up in Chapter 6. Furthermore, it includes experiments concerning the GAN-training procedure.

Chapter 8: Conclusion: is the final chapter of this thesis and contains a summary as well as thoughts about future work.

CHAPTER 2

Breast MR Imaging

This chapter is devoted to the description of the imaging data, to which the methodological steps in this thesis are applied. 3-dimensional images of the female breast of 20 high-risk patients for breast cancer were acquired using Magnetic Resonance Imaging (MRI). Each patient underwent the procedure of acquisition repeatedly in a time span ranging fro 2 to 13 years.

The aim of the following sections is to give an insight in the generation process of the images. The chapter is structured as follows: First, the physical background and basics of tomography are described in Sections 2.1 and 2.2. These sections are based on [21], if not stated differently. In the second part the focus lies on breast imaging in particular (Section 2.3).

2.1 Basics of Nuclear Spin

Nuclear spin is the physical property, on which MRI is based. Protons, electrons, and neutrons have a spin, acting like tiny magnetic gyroscopes with a magnetic moment $\vec{m}$. For MRI, the hydrogen nuclei, which are protons and are present in the human body, are used.

If the protons are brought in a constant magnetic field B in z-direction, they orientate themselves in two possible directions parallel to the z-axis, which are called (+) and (-) for simplicity. The two directions have different energy levels, the (+)-direction is energetically more favorable. Electromagnetic waves which deliver the energy difference lead to a change from the energetically favorable state to the more *expensive* one. This change of state is continuous. An exposure to the electro magnetic waves lasting for a time T_{90° causes the protons or their magnetic moment $\vec{m}$ to lie completely in a x-y-plane. If it lasts for $2T_{90^\circ}$ the z-component of $\vec{m}$ is reversed. When the exposure stops, the protons fall back into their initial position parallel to B. This is called relaxation.

Relaxation Times For MRI there are two relaxation times which have a significant meaning. The first one is the *spin-lattice relaxation* T_1. It describes how much time the protons need to reach their initial state after the exposure to the electromagnetic waves has stopped. The relaxation is caused through interaction of neighboring atoms.
The second relaxation time is the so called *spin-spin relaxation* T_2. Due to the magnetic field B, the protons are not motionlessly orientated parallel to the z-axis but precess around it. The additional electromagnetic waves increase the angle between the z-axis and the magnetic moment $\vec{m}$ and several protons precess in phase, which means that their transverse component (moment $\vec{m}$ projected on x-y-plane) points in the same direction. The time T_2 indicates how long the protons need to be out of phase. This is illustrated in Figure 2.1. The relaxation is also caused by atomic interactions. T_2 is always shorter or equal to T_1.

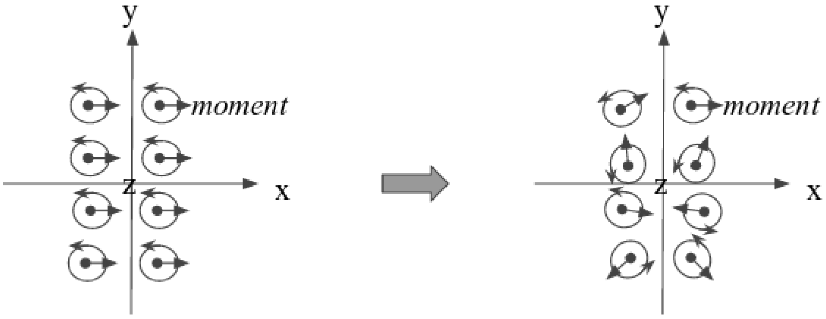

Figure 2.1: The figure illustrates the orientation of the traversal component of neighboring protons. Immediately after the electromagnetic impulse has stopped, all moments point in the same direction. After some time they are out of phase. Figure design is inspired from Figure 11.37 in [21]

2.2 Basics of Tomography

In MRI transverse magnetization is translated into an image. A human body is brought into a strong magnetic field in z-direction and through additional electromagnetic waves the magnetic moment of the hydrogen nuclei gets a transverse (x, y)-component. The transverse magnetization of a volume is the vectorial sum of the transverse components of magnetic moments in that volume divided by volume size. It is measured via an antenna, in which voltage is induced. The whole excited body volume contributes to the signal of the antenna. The task of tomography is therefore to encode the signals which correspond to different voxels in way so as to be able to reconstruct a 2-dimensional image (i.e. one layer of a 3D volume) from this encoding. For this purpose the following scheme is used additionally to the magnetic field in z-direction and the exposure to an electromagnetic pulse to bring the magnetic moments in x-y plane:

- Selective excitement using a magnetic gradient field in z-direction during electromagnetic excitement
- Encoding of phase using a gradient field in y-direction between electromagnetic excitement and readout by antenna
- Encoding of frequency using a gradient field in z-direction during readout by antenna

In the following the 3 steps are described in more detail. Besides the basic idea of extending the acquisition of one layer of the human body to the acquisition of a volume is mentioned.

Selective Excitement Selective excitement means that only one layer of the human body is excited so that its transverse magnetization is measured. This is done by engaging a gradient field in z-direction in addition to the main magnetic field and electromagnetic excitement. A gradient field in z-direction is a magnetic field, which is not constant in every direction but varies along the z-axis. It leads to a difference in precession frequency of the spins belonging to different z-planes and therefore to different energy levels needed to change the (+)- state into the (-)- state. The electromagnetic pulse only excites the layer with the correct precession frequency.
To ensure a strong signal the poles of the gradient field are inversed for half the time the electromagnetic stimulus lasted, after the latter has subsided. This enforces the magnetic moments in the excited layer to point in the same direction.

Encoding of Phase When all spins of the excited layer precess in x-y-plane pointing in the same direction, a gradient-field in y-direction, which is engaged for a certain time span, makes the spins along the y-axis point in different directions. The gradient is turned off before a signal is read out. The process of letting the spins precess in x-y-plane and engaging a gradient-field in y direction is repeated several times always with a different gradient in y-direction. This leads to different values of transverse magnetization measured by the antenna.

Encoding of Frequency Additional to selective excitement and encoding of phase, a gradient field in x-direction is engaged while the transverse magnetization is read out by the antenna for every gradient field for the encoding of phase. As a consequence, the information of voxel-placement is encoded by the frequency and each volume element sends its own unique signal. The antenna measures a mixture of frequencies, which is transformed into an image using the Fourier transform.

Echos and 3D Scanning The described way of translating transverse magnetization into images neglects the relaxation of spins after the electromagnetic pulse which brings the moments into x-y plane is turned off. It has to be turned off because otherwise the spins did not precess in x-y plane but around an axis whose angle with the z-axis increases more than 90°. Since the detected transverse magnetization gets weaker after the electromagnetic stimulus stops, there is not enough time for encoding of phase and frequency.

Therefore instead of a single pulse to bring the moments into x-y plane, a pulse, which rotates the orientation of the magnetic moments 180° is engaged after a pulse that rotates the moments 90° (in case of *Saturation Recovery*). This second pulse leads the spins to be in phase again and the antenna gets a signal. This signal is called *echo* and between the two pulses, there is enough time for the encoding of phase. The gradient for the encoding of frequency is engaged during the *echo*. For more information on *echos* the interested reader is referred to [21].
For the creation of 3-dimensional images one possibility is to acquire several layers successively.

2.3 Breast MRI

In breast imaging, MRI is an essential tool, whereas a special form of MRI, *Dynamic Contrast enhanced MRI* (DCE-MRI), is the most sensitive method for the detection of breast cancer [22]. Other variants of MRI used in breast imaging exist, as for example *Diffusion Weighted Imaging*. For an overview the interested reader is referred to [22], as this thesis focuses only on DCE-MRI imaging data.

DCE-MRI images are T1-weighted images. Different tissues are distinguished because of their different spin-lattice relaxation T_1. In order to obtain T_1- dependence, an impulse sequence with a short repetition time T_R of f.e. 200-500 ms is used. T_R is the time between an impulse, which rotates the moments of spins 90° and the 90°-impulse of the *Saturation Recovery* sequence (90°-impulse followed by 180°-impulse), mentioned in the last section. The time between the 90°-impulse and 180°-impulse has also to be short, lasting f.e. for 7.5-15 ms.
Fatty tissue appears brighter in T1-weighted images than glandular tissue, which is visible in Figure 2.2 on the left-hand side. It is also possible to suppress the signal of fatty tissue, which leads corresponding regions to appear black in the image. This is visible in Figure 2.2 on the right-hand side.

The term "dynamic contrast enhanced" refers to the fact that a contrast agent is injected into the veins of the patients prior to scanning so as to enhance the contrast of lesions compared to healthy tissue. Tumors, especially aggressive ones, create a vasculature with abnormal vessel permeability to support their increasing demand for oxygen and nutrients [22]. This abnormal vasculature and permeability is depicted by DCE-MRI through the assessment of kinetic enhancement after the injection of the contrast agent [22]. The analysis includes the acquisition of at least 2-3 post-contrast images after the injection [22], then an image acquired before the injection is subtracted from these images. Lesions are expected to appear brighter than healthy tissue in the difference images [33]. In Figure 2.3, there is an example given.

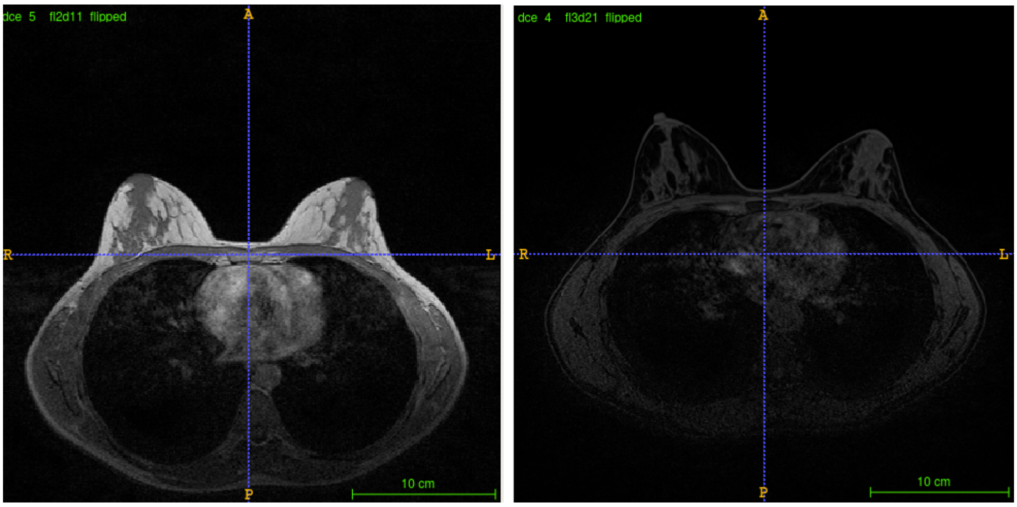

Figure 2.2: Examples of DCE-MR images with (left) and without fat suppression (right). The images are part of a data set received from the General Hospital of Vienna.

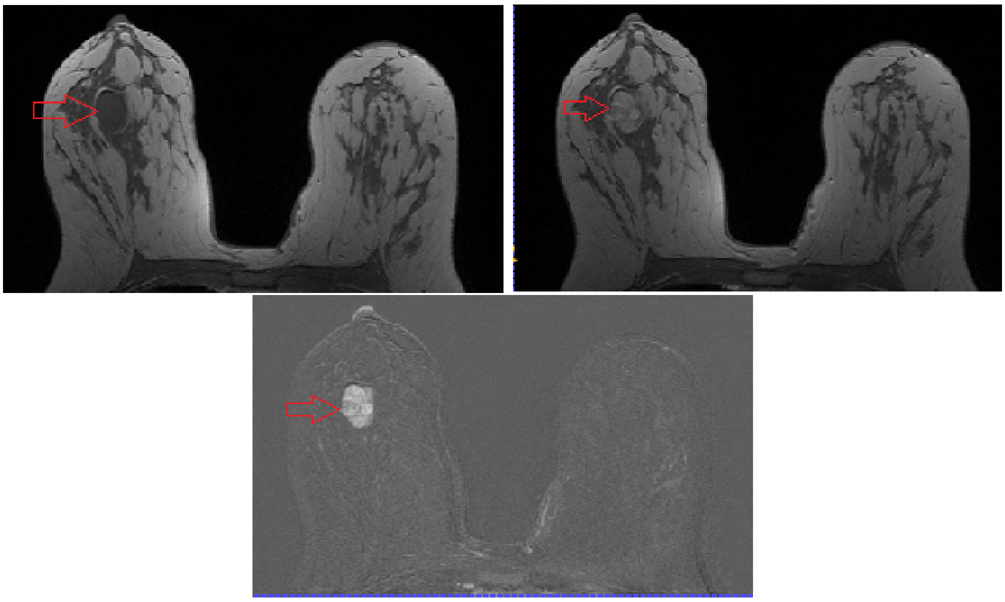

Figure 2.3: Examples of DCE-MRI images: before contrast agent injection (top,left), first image acquired after injection (top, right), and difference image, where a lesion is clearly visible (bottom) The images are part of a data set received from the General Hospital of Vienna.

To distinguish between malignant and benign lesions kinetic enhancement curves are used [22]. Lesions which continue to enhance over the entire acquisition period or where the signal gain is slowed down in the late post-contrast phase are classified as benign [33]. Malign lesions are more likely to show a plateau after an early increase in intensity or the signal decreases immediately after the intensity peak [33]. This is an approach based on the question, what happens after the initial signal increase. Malignant lesions are more likely to have a high signal intensity in the first post-contrast image than benign ones [33].

2.4 Discussion

In this chapter the physical background of MRI has been discussed as well as its application in breast imaging. Through MRI the magnetization of protons in the human body is measured. For breast imaging, the magnetization measured of a specific tissue is distinguished from another tissue type based on different T1 times. This leads to T1-weighted images.

Furthermore, a contrast agent is injected after the acquisition of a reference image to asses the enhancement kinetics of the breast tissue. This also includes the acquisition of post-contrast images. For distinction of malignant and benign lesions kinetic enhancement curves are used.

CHAPTER 3

Image Registration

In this thesis registration is used to reach spatial correspondence of the images belonging to the same patient but not to the same point in time. The focus of this chapter is therefore on non-rigid registration as well as on approaches, which are able to align images acquired by different modalities.
Image registration, especially non-rigid registration, is one of the main challenges in medical image analysis [52]. A full discussion of all registration approaches described in papers published in the last decade is far beyond this thesis and the interested reader may refer to [52] for bibliographic overview. This chapter discusses only some examples. In the following, an overview of a registration framework is given. Then examples of its components are described in more detail.

3.1 Constituents of a Registration Framework

As described in [52], the aim of registration is to align two images, a source image $\boldsymbol{S}$ and a target image $\boldsymbol{T}$. These images are in the image domain Ω and related by the transformation $\boldsymbol{W}$. The optimal transformation for alignment is found by optimizing a function of the form

$$\mathcal{M}(\boldsymbol{T}, \boldsymbol{S} \circ \boldsymbol{W}) + \mathcal{R}(\boldsymbol{W}). \tag{3.1}$$

$\mathcal{M}$ describes how well the alignment between $\boldsymbol{S}$ and $\boldsymbol{T}$ is and it is referred to as distance measure, (dis)similarity measure or matching criterion. $\mathcal{R}$ is a regularization term in which prior knowledge or desirable quantities of the transformation are included.
A registration framework consists of three components, which are discussed separately in the following sections. Those are

- A matching criterion

- A deformation model
- An optimization strategy

3.2 Examples of Matching Criteria: Mutual Information and Cross-Correlation

This section refers to the dissimilarity measure term in (3.1) and discusses two examples of matching criteria, namely *Mutual Information* and *Cross-Correlation*.

3.2.1 Mutual Information (MI)

MI is an information theoretic measure used for inter-modal image registration [52]. It was independently developed by Collignon *et al.* [30] and Viola and Wells [56] and led to further research interest in that topic with an overview found in [47]. MI is defined based on the *Shannon Entropy* H [12],

$$H = -\sum_{i}^{m} p_i \log p_i, \tag{3.2}$$

where $p_i, i = 1, ...m$ is the probability of the occurrence of an event e_i. For images this events are the gray values in the image. The probability is estimated by counting how often a gray value occurs and dividing that number by the total number of voxels.
To adapt the *Shannon Entropy* to image registration an idea of Hills *et al.* [13] was used by Collignon *et al.* [29] and Studholme *et al.* [19]. This idea is to use a *joint histogram*, which shows all combinations of gray values between two images $\boldsymbol{A}$ and $\boldsymbol{B}$ and how often they occur for corresponding points in them. Based on the joint distribution of gray values the joint *Shannon Entropy* is then given as

$$H_{\boldsymbol{AB}} = -\sum_{i,j} p(i,j) \log p(i,j). \tag{3.3}$$

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are well aligned the joint entropy is minimized. In *Mutual Information* the entropies of $\boldsymbol{A}$ and $\boldsymbol{B}$ are included additionally, yielding the expression

$$MI(\boldsymbol{A}, \boldsymbol{B}) = H_{\boldsymbol{A}} + H_{\boldsymbol{B}} - H_{\boldsymbol{AB}} \tag{3.4}$$

The *Shannon entropy* takes only voxels into account which overlap. This leads to a minimized entropy when just parts of the background are overlapping and therefore taken into account. This would lead to trivial but wrong registration objective function minima. MI tackles this problem by including the marginalized entropies of the images. They are larger, if an anatomical structure is included and not only background. Optimizing the mutual information of two images, therefore means including as much as possible of the anatomical structure and minimizing the joint entropy of the images in order to align them well. A normalized version also exists and is proposed by Studholme *et al.* [20],

$$NMI(\boldsymbol{A}, \boldsymbol{B}) = \frac{H_{\boldsymbol{A}} + H_{\boldsymbol{B}}}{H_{\boldsymbol{AB}}}. \tag{3.5}$$

It is more robust if the overlap is small and the relative areas of background and structure even out [47].

3.2.2 Cross-Correlation

Cross-Correlalation (CC), also called *Correlation Coefficient* or *Normalized Correlation* is an intensity-based based matching criterion and assumes an linear relationship between the signal intensities of two images $\boldsymbol{A}$ and $\boldsymbol{B}$ [52]. If the gray values of the images are presented as vectors of length n with entries A_i and B_i, respectively, then the *Cross-Correlation* among the images is

$$CC(\boldsymbol{A}, \boldsymbol{B}) = \frac{\sum_i^n (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_i^n (A_i - \bar{A})^2 \sum_i^n (B_i - \bar{B})^2}}, \tag{3.6}$$

with

$$\bar{A} = \frac{1}{n} \sum_i^n A_i \quad \text{and} \quad \bar{B} = \frac{1}{n} \sum_i^n B_i \tag{3.7}$$

CC as defined above is sensitive to outliers. Therefore Kim and Fessler [35] developed in 2004 a more robust variant. Avants *et al.* [42] compute a local *Crosscorrelation*, which is computed of a window centered around a point x in the image according to the above equations (3.6) and (3.7).

3.3 Examples of Deformation Models: SyN and Demons

The choice of a transformation model reflects, which kinds of transformations are acceptable in order to align two images [52]. The degrees of freedom of the model correspond to the parameters, which are optimized by the optimization strategy and range from six degrees of freedom to millions [52]. An *affine transformation* for example has 12 degrees of freedom and is given by

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \tag{3.8}$$

where the point $\boldsymbol{x}$ is transformed by multiplication with matrix $\boldsymbol{A}$ and adding vector $\boldsymbol{b}$. In the following two transformation models ae discussed in more detail. The first one is *Symmetric Normalization* (SyN) and the other one corresponds to a class which called *Demons* approaches.

3.3.1 Symmetric Normalization

SyN is proposed by Avants *et al.* [42] and belongs to a framework known as large deformation diffeomorphic metric mapping [52]. In general this framework provides diffeomorphic transformations [52]. A diffeomorphism is an invertible mappping, where the forward and the backward mapping are differentiable.
As described in [42], solutions for the transformation for registration are restricted to the diffeomorphic space $Diff_0$ with homogeneous boundary conditions [42]. For a diffeomorphism ϕ of domain Ω, the transformation of an image $\boldsymbol{I}$ into a new coordinate system is given by $\phi\boldsymbol{I} = \boldsymbol{I}(\phi(\boldsymbol{x}, t = 1))$. This transformation is parameterized by time t, the spatial coordinate $\boldsymbol{x}$ and a velocity field $\boldsymbol{v}(\boldsymbol{x}, t)$ on Ω through

$$\phi(\boldsymbol{x}, 1) = \phi(\boldsymbol{x}, 0) + \int_0^1 \boldsymbol{v}(\phi(\boldsymbol{x}, t), t)dt, \tag{3.9}$$

where $\boldsymbol{v}$ is indexed at $\phi(\boldsymbol{x}, t) = \boldsymbol{y}$. Based on this notation a distance D between two images is defined as

$$D(\phi(\boldsymbol{x}, 0), \phi(\boldsymbol{x}, 1)) = \int_0^1 ||\boldsymbol{v}(\boldsymbol{x}, t)||_D dt, \tag{3.10}$$

where $||f||_D = ||Df||_{L_2}$, with a linear differential operator D.

Avants *et al.* [42] use the property, that the diffeomorphism ϕ is allowed to be decomposed into two parts ϕ_1 and ϕ_2. This is used to divide the transformation path from an image $\boldsymbol{I}$ to an image $\boldsymbol{J}$ into two halves. The deformation is divided between the images and a symmetric formulation of the optimization problem for image registration is

$$\begin{aligned} E_{SyN} = \inf_{\phi_1} \inf_{\phi_2} \int_{t=0}^{t=1} \{||\boldsymbol{v}_1(\boldsymbol{x}, t)||_D^2 + ||\boldsymbol{v}_1(\boldsymbol{x}, t)||_D^2\}dt \\ + \int_\Omega |\boldsymbol{I}(\phi_1(0.5)) - \boldsymbol{J}(\phi_2(0.5))d|\Omega. \end{aligned} \tag{3.11}$$

Subject to each $\phi_i \in Diff_0$ which fulfills

$$\frac{d\phi_i}{dt} = \boldsymbol{v}_i(\phi_i(\boldsymbol{x}, t), t) \tag{3.12}$$

$$\phi_i(\boldsymbol{x}, 0) = Id \tag{3.13}$$

$$\phi_i^{-1}(\phi_i) = \phi_i(\phi_i^{-1}) = Id. \tag{3.14}$$

The solution to the above optimization problem (3.11)-(3.14) provides the *Symmetric Normalization* solution.

3.3.2 Demons

In 1998 Thirion [43] proposed an image matching technique based on a diffusion process. His algorithm he called *Demons* algorithm. The concept of *Demons* is adapted from thermodynamics, where a gas mixture consisting of two types of particles a and b is assumed. This mixture is filled into two compartments A and B, which are separated by a semi-permeable membrane. This membrane contains *Demons* which are able to distinguish between particles of type a and b. Besides it allows only a-particles to diffuse into compartment A and particles of type b to diffuse into compartment B. This leads to compartment A being filled only with particles of type a and compartment B with b-particles, as shown in Figure 3.1.

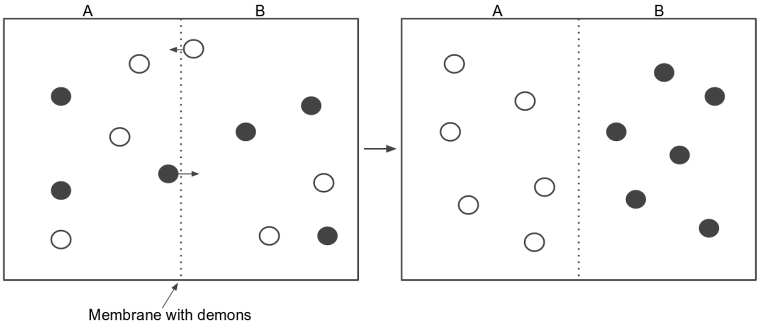

Figure 3.1: The image presents the diffusion process between two compartments, separated by a semi-permeable membrane. Both compartments are filled with a gas mixture consisting of two compounds, where compound a is only allowed by *Demons* to diffuse into compartment A but not into B. The same yields for b but vice-versa. Figure adapted from [43]

So as to extend the idea of *Demons* to image matching, it is assumed that the contour of an object O in the target image $\boldsymbol{T}$ is a membrane, with demons scattered along it. To each point, where a *Demon* sits, a force perpendicular to the contour is applied, pointing either inside or outside, depending on the labeling of the point of the source image they are looking at. Points of the source image are labeled "inside" if they have to be pushed inside the object O and "outside", if they have to be pushed out of it. The force mentioned, is given by Thirion [43] as

$$\boldsymbol{v} = \frac{(s-t)\nabla t}{(\nabla t)^2 + (s-t)^2} \tag{3.15}$$

where s is the intensity in $\boldsymbol{S}$ at a point P and t the intensity in $\boldsymbol{T}$. This force enables the behavior of a *Demon*, since it labels implicitly points of the source image as "inside" or "outside" based on its intensity compared to the corresponding intensity of the target image and sorts correspondingly.

To calculate the transformation $\mathcal{T}$ to align $\boldsymbol{S}$ to $\boldsymbol{T}$ an iterative scheme is used, which all demon approaches share [52]. It includes

- Calculating for each *Demon* the associated force $\boldsymbol{v}$

- Computing transformation $\mathcal{T}_{i+1}$ from $\mathcal{T}_i$ and the demon forces

An illustration of the working principle of *Demons* is given for two gray discs in Figure 3.2.

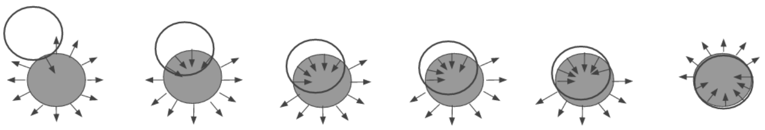

Figure 3.2: The figure illustrates the principle of operation of demons. The upper disc is aligned with the bottom disc, where the *Demons* on the bottom disc keep background pixels outside and let only pixels of the disc inside. Figure adapted from [43]

For 3D medical image registration it has been proven useful to select all pixels of $\boldsymbol{T}$ as *Demons* and not only contour points [43]. Further successful variants of the proposed iterative scheme included, assuming the space of transformations to be free form and applying Gaussian filters to the update displacement field $\boldsymbol{u}_{i_j}$ during each iteration or to the total displacement field after each iteration $\boldsymbol{u_i}$ [52] with,

$$\mathcal{T}_i(\boldsymbol{S}) = \boldsymbol{S} + \sum_{m=1}^{j} \boldsymbol{u}_{i_m}(\boldsymbol{S}) = \boldsymbol{S} + \boldsymbol{u}_i(\boldsymbol{S}) \tag{3.16}$$

describing a transformation by adding a displacement field to the original image frame.

Thirion's *Demons* approach [43] was developed further, including the introduction of *Diffeomorphic Demons* by Vercauteren *et al.* [50] and the extension to other similarity measures, such as *Mutual Information* [61]. The original *Demons* algorithm is only applicable to intra-modality image registration, because its formulation relies on the hypothesis that corresponding structures have the same intensity values. Incorporating MI in the *Demons* approach, allows also its application to inter-modality registration.

A particular example of a *Demons* algorithm is *Ezys* [55], which belongs to the class of *Diffeomorphic Demons*. It is implemented to run on Graphical Processing Units (GPUs) and allows the use of *Normalized Mutual Information* or *Cross-Corleation* as matching criteria. It also uses a Gaussian filter to regularize the update displacement fields $\boldsymbol{u}_{i_j}$ but an anisotropic filter applied to the full displacement field $\boldsymbol{u}_i$. This kernel incorporates knowledge about the likelihood of a point in the image to belong to a surface. If a point is likely to lie on a surface less smoothing is applied to it, to prevent boundaries from getting blurry. Surface detection is performed through an interactive process in which weights are assigned to each position in the image. High weights indicate a surface. The detection is done after a noise reduction step.

3.4 Optimization Strategy

The aim of optimization is to find the optimal transformation to align two images. This is done by maximizing or minimizing an objective function, as given in (3.1). Several optimization strategies exist, including *continuous optimization*, discrete *optimization, greedy approaches* and *evolutionary algorithms* [52]. The choice of the best strategy depends on the properties of the optimization problem. *Discrete optimization* solves problems where variables assume discrete value and *continuous optimization* is used when real values are assumed [52]. In the latter case, gradient based optimization procedures are used, such as *Stochastic Gradient Descent* [52], which is described in Chapter 4. *Greedy approaches* make at each step a locally optimal choice out of a set of possible solutions. It is a good choice for feature driven registration [52]. *Evolutionary approaches* have mainly been used for linear registration and are inspired by the theory of evolution [52]. More information on these strategies can be found in [52].
Further strategies are the use of a *coarse-to-fine scheme* and a *multi-step approach* [63]. The idea of a *coarse-to-fine scheme* is to perform the registration first on a coarse scale, where a fewer pixels are involved in registration. This spatial mapping is then used to initialize the registration at the next finer scale. This strategy improves speed, accuracy, and robustness [63]. The *coarse-to fine scheme* is expendable by a *multi-stage approach*, where, for example, an affine transformation at the first resolution step is followed by a deformable transformation at the second resolution step. It is also possible to apply the same transformation type to more than 1 resolution level [52].

3.5 Discussion

This chapter has been concerned with giving a quick overview on image registration. The main components of the registration process have been identified and examples for them have been given. Those compounds are a matching criteria, a transformation model and an optimization strategy. *Mutual Information* and *Crosscorrelation* haven been described as examples for matching criteria and *Symmetric Normalization* and *Demons* for the transformation model. Finally, examples of optimization strategies have been mentioned. The registration procedure used in this thesis includes a two-step approach, where an affine transformation is performed prior to a nonrigid one. Further, each kind of transformation is calculated using a *coarse-to-fine scheme.*

CHAPTER 4

Generative Adversarial Networks

Generative Adversarial Networks(GANs) were invented by Goodfellow et al.[70] in 2014. They are a framework consisting of two models, which are trained in parallel in an adversarial process [70]. In this thesis these models are so-called *Deep Networks*. The latter correspond to a mathematical interpretation of a collection of neurons in a brain, which are interconnected to each other [32].
There are neurons, which receive an input. This input is then transformed in some way, when passed from one neuron to others through their connections and finally there is some kind of output. The connections of the neurons have to be formed in a way, so as to receive the desired output for an input. Therefore training of *Deep Networks* is required. The neuronal structure of *Deep Networks* leads in general to nonlinear functions [32]. Through training, this functions are created to fulfill certain tasks, for example learning to reproduce the behavior of the XOR-function, a task which cannot be done through a single linear function [32].

This chapter describes the idea behind GANs and how this framework can be applied to images, the data source on which this thesis is based. The *Deep Convolutional Generative Adversarial Network* (DCGAN) is a variant of GANs, which is designed for this purpose [60]. Starting from a closer look at adversarial training of two *Deep Networks* in Section 4.1, they themselves are described in more detail concerning their architecture, design and optimization in Section 4.2. Subsequently, Section 4.3 deals with DCGANs and how they can serve as anomaly detectors as we do in this thesis. Finally, Section 4.4 points out ways to improve the GAN-training procedure.

4.1 The idea behind Generative Adversarial Networks

In this section the idea of GANs is described in more detail. It is mainly based on the paper of Goodfellow *et al.* [70].

GANs correspond to a framework consisting of two models. The first one is a generative model and is called *Generator* G. Its task is to learn the distribution of the training data to be able to generate samples looking like the original data used for training. The second model is the *Discriminator* D, which is a discriminative model and has to learn to differentiate between samples of the distribution of the training set and examples produced by G. The models are trained in an adversarial process to improve their abilities. A descriptive example of this adversarial process is given in [70], where the *Generator* corresponds to a counterfeiter, who wants to produce fake money, which is indistinguishable from real money. His opponent is the police, who wants to recognize the counterfeiter's money as fake currency and is represented by D. Over time the counterfeiter improves his or her techniques and so does the police until the fake money cannot be distinguished from real currency. In the following, this process is formalized. The *Generator* G and the *Discriminator* D are both implemented by *Deep Networks*. The latter are more precisely described in the next section, for this part it is sufficient to think of G and D as nonlinear functions.

Let p_g be the distribution of G over data $\boldsymbol{x}$ and $p_z(\boldsymbol{z})$ a prior of input noise variables. The mapping to data space is defined by $G(\boldsymbol{z}, \boldsymbol{\theta}_g)$, which means G maps from the input noise to the data, based on some parameters $\boldsymbol{\theta}_g$ subject to optimization. Further, a mapping $D(\boldsymbol{x}, \boldsymbol{\theta}_d)$ from data space to a scalar-valued output is defined. This is the *Discriminator*, assigning a probability that $\boldsymbol{x}$ came from the data rather than from p_g. For training, the two models play a min-max game against each other, in which D maximizes the probability to assign the correct label to data sampled from p_g and the training data. G is trained to minimize $\log(1 - D(G(\boldsymbol{z})))$, which is formalized, together with the aim of D, in the following expression with value function $V(G, D)$

$$\min_G \max_D V(G, D) = \min_G \max_D \{\mathbb{E}_{\boldsymbol{x} \sim p_{data}}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_z}[\log(1 - D(G(\boldsymbol{z})))]\}. \tag{4.1}$$

V is maximal in D, if it assigns 1 to x and 0 to $D(G(\boldsymbol{x}))$, which corresponds to labeling training data as "real" and data produced by G as "fake". Besides the value function is minimal in G, if it is able to fool D, meaning that D labels a generated example as coming from the data distribution.

Theoretical results in [70] show, that the V reaches is optimum, if $p_g = p_{data}$. If D is allowed to reach its optimum for a fixed G and then G is updated according to (4.1), p_g becomes equal to p_{data} and convergence is reached.
Figure 4.1 gives a descriptive overview of the training procedure. The distribution of D (blue, dashed line) is simultaneously updated to discriminate the distribution of G (green, solid line) from the distribution of the training data (black dotted line) The z-line and x- line represent the random input noise and the space of data x coming from the true distribution or p_g The black arrows symbolize the transformation from input noise to data samples produced by the *Generator*.

(a) A pair of G and D is considered. The *Generator*'s distribution p_g is close to the true distribution of the data p_{data}. D's distribution shows regions of rather high density corresponding regions with high densities of real data and it shows rather low density for regions, where p_g is high. This makes D a partially good classifier.

(b) D is trained k-times so as to converge to the optimal discriminator for the actual generator.

(c) G is updated based on the gradient of the *Discriminator* and moves its distribution closer to examples which are more likely to be classified as data.

(d) When convergence is reached p_g is identical to p_{data} and D is not able to distinguish between those two distributions, leading to a probability of $\frac{1}{2}$ of being classified as training or generated data for any sample.

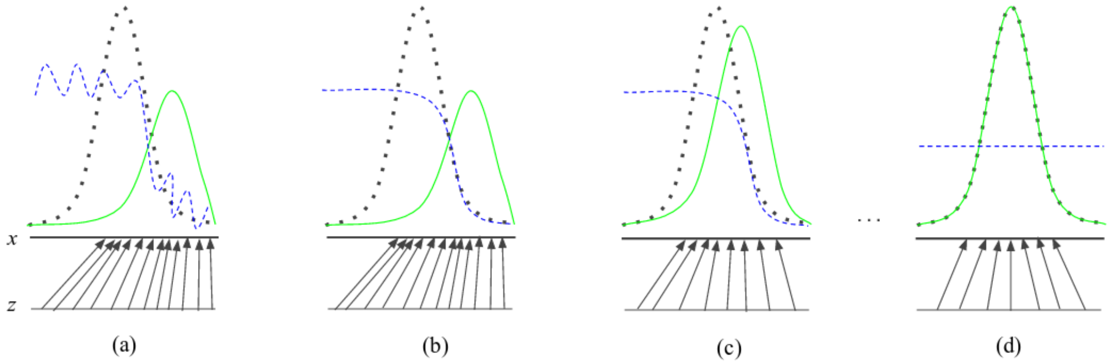

Figure 4.1: The image presents an overview of the training procedure of *Generator* and *Discriminator*. Through training the *Generator*'s distribution p_g (green) is moved towards the data distribution p_{data} (black) based on the gradient of D. Updates of D and G are alternating. Figure adapted from [70]

In practice, the training of G and D is implemented as an iterative process, where optimization of parameters is done by alternating between k steps of updating the *Discriminator* and one step of updating the *Generator*. Pseudo-code for training a GAN is given in Algorithm 1.

Algorithm 1 Training Generative Adversarial Networks using Minibatch Stochastic Gradient Descent, The algorithm is taken from [70]

```
for number of training iterations do
    for $k$ steps do
        Sample minibatch of $m$ noise samples $\{\boldsymbol{z}^{(1)}, ..., \boldsymbol{z}^{(m)}\}$ from noise prior $p_z$).
        Sample minibatch of $m$ examples $\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(m)}\}$ from $p_{data}$.
        Update the Discriminator by ascending its stochastic gradient:
            $$\nabla_{\boldsymbol{\theta}_d} \frac{1}{m} \sum_{i=1}^{m} [\log D(\boldsymbol{x}^{(i)}) + \log(1 - D(G(\boldsymbol{z}^{(i)})))].$$
    end for
    Sample minibatch $m$ noise samples $\{\boldsymbol{z}^{(1)}, ..., \boldsymbol{z}^{(m)}\}$ from noise prior $p_z$
    Update the Generator by ascending its stochastic gradient
        $$\nabla_{\boldsymbol{\theta}_g} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(\boldsymbol{z}^{(i)}))).$$
end for
```

Minibatch Stochastic Gradient Descent and *Backprobagation* are used to calculate the updated parameters of G and D. The first one is a gradient-based optimization algorithm used in machine learning problems and the latter is a procedure to compute the gradients of *Deep Networks* [32]. Both are described later in this chapter.

4.2 Deep Networks

Deep Networks are inspired by neurons of the human brain and the way information is processed by them [32]. There are different variants including the *Feedworward Neural Network* (FNN) and the *Convolutional Neural Network* (CNN) [32]. Both are described in this section in terms of their design and ways to train them to get the insight in *Deep Networks* needed for this thesis. For more information on *Deep Networks* the interested reader may refer to [32], which is the main source for this part.

4.2.1 Design of Feedforward Networks

Feedforward Neuronal Networks are used to approximate a function f^* to fulfill a certain task, whereas information flows from an input $\boldsymbol{x}$ though intermediate computation to the output $\boldsymbol{y}$ without any feedback. Thus they are called "feedforward". Besides, they correspond to a network, because they are composed of many different functions. Thereby, the architectural structure of FNNs consists of layers. For example, a chain structure $\boldsymbol{y} = f^3(f^2(f^1(\boldsymbol{x})))$ with the functions f^1, f^2 and f^3, has f^1 as first layer, f^2 as second layer and f^3 as third layer.

The last layer of the network is also called *output layer* and the overall chain length of the network is called *depth.* During training $f(\boldsymbol{x}) = \boldsymbol{y}$ is driven to approximate $f^*(\boldsymbol{x})$, where the input data $\boldsymbol{x}$ provides noisy examples, for which the desired output $\boldsymbol{y}$ is known presented by assigning the label $\boldsymbol{y} \approx f(\boldsymbol{x})$. Therefore, only the *output layer* is expected to produce a predefined output, whereas the behavior of all other layers is not directly specified. The algorithm itself has to learn how to use those layers in order to approximate f^*. Those layers are called *hidden layers*, since their output is not specified by the training data.

Each *hidden layer* is typically vector valued and the number of components or *units* of this vectors determines the *width* of the model. The components can also be thought of as neurons working in parallel, receiving input from other neurons and learning its own activation rule. The choice of functions f^i which are used to compute the vector valued representations are also loosely inspired from neuroscience and represent the connections between neurons. Examples of them used in *output layers* and *hidden layers* are given in the following.

Output layers For this part it is assumed, that hidden features $\boldsymbol{h} = f(\boldsymbol{x}, \boldsymbol{\theta})$ are given, which have been computed based on the input $\boldsymbol{x}$ and some parameters $\boldsymbol{\theta}$. The latter are updated during training so as to get an approximation of the desired f^*.

The first design possibility are Linear Units, which are based on an affine transformation. For given features $\boldsymbol{h}$ the output $\hat{\boldsymbol{y}}$ is produced by

$$\hat{\boldsymbol{y}} = \boldsymbol{W}^T \boldsymbol{h} + \boldsymbol{b}, \tag{4.2}$$

with the matrix $\boldsymbol{W}$ and the translation-vector $\boldsymbol{b}$. Linear units are used to model the mean of a conditional Gaussian distribution $p(\boldsymbol{y}|\boldsymbol{x})$. Besides their gradients do not saturate, which is advantageous for gradient-based optimization algorithms, such as *Stochastic Gradient Descent.*

Next Sigmoid Units are discussed. They are used to predict the value of a binary variable $\boldsymbol{y}$, as it is the case for a classification problem with two classes. This type of unit is defined by

$$\hat{\boldsymbol{y}} = \sigma(\boldsymbol{w}^T \boldsymbol{h} + b), \tag{4.3}$$

where σ is the logistic sigmoid [32] applied to a linear layer $z = \boldsymbol{w}^T \boldsymbol{h} + b$ and converts z into a probability. Using the sigmoid is motivated by first constructing an unnormalized probability distribution $\tilde{P}(y)$ and assuming, that the unnormalized log probabilities are linear in z and y,

$$\log \tilde{P}(y) = yz. \tag{4.4}$$

Exponentiating and normalizing the probabilities

$$\tilde{P} = \exp yz, \tag{4.5}$$

$$P(y) = \frac{\exp yz}{\sum_{y'=0}^{1} \exp y'z} \tag{4.6}$$

yields a Bernoulli distribution controlled by a sigmoid transformation of z

$$P(y) = \sigma((2y-1)z), \tag{4.7}$$

since

$$\sigma(z) = \frac{1}{1+\exp^{-z}}. \tag{4.8}$$

The z variable which defines such a distribution over binary variable is called *logit*.

Thirdly, Softmax Units are described. They are used, if one aims to present a probability distribution over a discrete variable with n possible values. This is for example the case, if a classifier, which represents the probability of n different classes, is desired. For binary out put variables, the production of a single number

$$\hat{y} = P(y = 1|\boldsymbol{x}) \tag{4.9}$$

was wished. In order to generalize to a variable with n values, a vector $\hat{\boldsymbol{y}}$, with $\hat{y_i} = P(y = i|\boldsymbol{x})$ needs to be produced. This includes $\hat{y_i}$ to lie between 0 and 1, as well as the entire vector summing to 1. This approach generalizes to the multinoulli distribution.
A linear layer

$$\boldsymbol{z} = \boldsymbol{W}^T\boldsymbol{h} + \boldsymbol{b} \tag{4.10}$$

is first applied to features $\boldsymbol{h}$ followed by the softmax-function,

$$softmax(\boldsymbol{z}_i) = \frac{\exp(z_i)}{\sum_j \exp(z_i)}. \tag{4.11}$$

From a neuroscientific point of view the softmax-function creates competition between the neurons, which participate in it. Its outputs always sum two one, which means that an increase in one value leads to a decrease in the other values. This is comparable to the lateral inhibition, which is believed to exist between nearby neurons in the cortex [32].

Hidden Layers The design of *hidden units* is a process of empirical experimentation. Some kinds of *hidden units* are not continuously differentiable having some points at which the derivative is not defined. This is a theoretical problem for gradient based optimization. However, these functions have a defined left and a defined right derivative, which is sufficient in practice. [32]
Most *hidden units* are described by an affine transformation $\boldsymbol{z} = \boldsymbol{W}^T + \boldsymbol{b}$ and distinguished by an element-wise nonlinear function $\boldsymbol{h} = g(\boldsymbol{z})$ following the affine transformation with features $\boldsymbol{h}$.

Rectified Linear Units are defined by the function

$$g(\boldsymbol{z}) = \max(0, z) \tag{4.12}$$

and are similar to linear units, except that they are 0 for half of the domain. for $z > 0$ the unit is active and the gradient is 1. It is 0 for $z < 0$.
For gradient based optimization algorithms, a gradient of 0 does not provide information on how to update parameters, which are subject to training. Several generalizations of the rectified linear unit exist, so that there is a nonzero gradient almost everywhere. These use a nonzero slope α_i, when $z_i < 0$:

$$h_i = g(\boldsymbol{z}, \boldsymbol{\alpha})_i = max(0, z_i) + \alpha_i \min(0, z_i) \tag{4.13}$$

Examples are given bellow.

- Absolute value rectification : It fixes $\alpha_i = -1$ to obtain $g(z) = |z|$ and is used for object recognition from images [71].
- Leaky ReLu: It fixes α_i to a small value like 0.01 [8].
- parametric Relu: It treats α_i as learnable parameter [40].

Maxout Units [69] generalize rectified linear units even further, by dividing $\boldsymbol{z}$ into groups of k values. Each maxout unit then outputs the maximum of one of this groups

$$g(\boldsymbol{z})_i = \max_{j \in \mathbb{G}^{(i)}} z_j, \tag{4.14}$$

where $\mathbb{G}^{(i)}$ is the set of indices of the input values belonging to group i, $\{(i-1)k+1, ..., ik\}$. This unit allows to learn a piecewise linear function, which responds to multiple directions in the input $\boldsymbol{x}$ space. In Goodfellow *et al.* [69] it is used for classification tasks on 4 different datasets.

The Hyperbolic Tangent and Logistic Sigmoid have been used before the introduction of rectified linear units,

$$g(z) = \tanh(z) \quad \text{and} \quad g(z) = \sigma(z). \tag{4.15}$$

The use of the logistic sigmoid of as *hidden unit* is now widely discouraged, whereas the hyperbolic magnets is still used [32].

Beside the already mentioned functions, some of the described *output units* are also used as *hidden units*. Examples are the softmax unit and linear units [32].

Other design considerations are the width and depth of the network. Besides the choice of the cost function used for training the network is coupled with the choice of the output units. Most *Deep Networks* use maximum likelihood, which means the cost function is the negative log-likelihood and of the form

$$J(\boldsymbol{\theta}) = -\mathbb{E}_{\boldsymbol{x},\boldsymbol{y} \sim \hat{p}_{data}} \log p_{model(\boldsymbol{y}|\boldsymbol{x})}, \tag{4.16}$$

with the empirical distribution $\hat{p}_{data}$. An advantage is that a specified model $p(\boldsymbol{y}|\boldsymbol{x})$ automatically gives a cost function $\log p(\boldsymbol{y}|\boldsymbol{x})$. Further, for gradient based learning, the gradient of the cost function must not become 0 to guide the algorithm well, because a gradient of zero gives no information on the direction in which the algorithm should move [32]. If the output units saturate, the cost function also does. The negative log-likelihood helps to overcome this problem for some cases, like the sigmoid units and softmax units which involve the exp-function.
Regularization of this cost function is also possible. The interested reader may refer to [32] to read an overview of regularization techniques.

4.2.2 Design of Convolutional Networks

Convolutional Neural Networks (CNNs) are specific *Deep Networks*, which are designed to process data with a grid-like topology, such as digital images. The convolution operation forms the basis of this kind of network. The discrete convolution is defined as

$$\boldsymbol{s}(t) = (\boldsymbol{x} * \boldsymbol{w})(t) = \sum_{a=-\infty}^{\infty} \boldsymbol{x}(a)\boldsymbol{w}(t-a) \tag{4.17}$$

with the *input* $\boldsymbol{x}$ being a multidimensional array of data and the *kernel* $\boldsymbol{w}$ a multidimensional array of parameters that are adapted during training. $\boldsymbol{s}$ is referred to as *feature map*. The multidimensional arrays are referred to a tensors with zero entries almost everywhere. In practice, the infinite sum is therefore calculated as summation over a finite number of values.
To convolve over two axes, the formula

$$\begin{aligned} \boldsymbol{S}(i,j) = (\boldsymbol{I} * \boldsymbol{K})(i,j) &= \sum_m \sum_n \boldsymbol{I}(m,n)\boldsymbol{K}(i-m,j-n) && (4.18)\\ &= \sum_m \sum_n \boldsymbol{I}(i-m,j-n)\boldsymbol{K}(m,n) && (4.19) \end{aligned}$$

is used, whereas the second equality is more straight forward to implement. A related function called *cross-correlation* is also implemented as convolution and given by

$$\boldsymbol{S}(i,j) = (\boldsymbol{I} * \boldsymbol{K})(i,j) = \sum_m \sum_n \boldsymbol{I}(i+m,j+n)\boldsymbol{K}(m,n). \tag{4.20}$$

In Figure 4.2, there is an illustrative example of how convolution works.

Discrete convolution corresponds to a multiplication by a matrix, with several entries constrained to be equal to other entries. Besides, it is a sparse matrix, since the kernel is usually smaller that the input image I. Convolution allows sparse interactions, parameter sharing and equivariant representations. Those ideas are described in more detail bellow.

Sparse Interactions CNNs make use of sparse interaction, which means that not every input unit interacts with every output unit in contrast to *fully connected layers*. This is accomplished by choosing the kernel to be smaller than the input image. As a

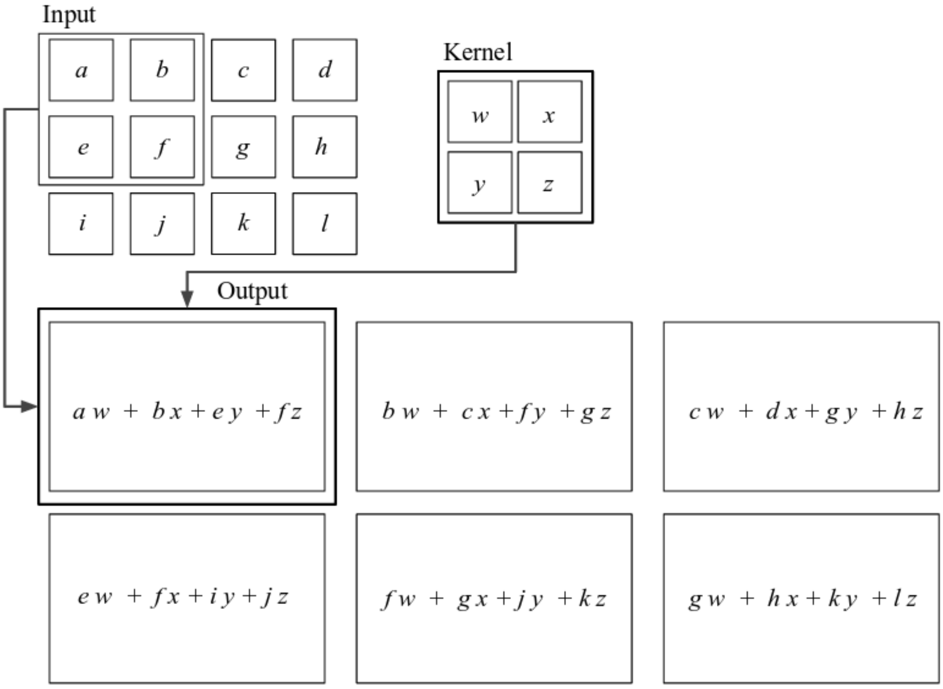

Figure 4.2: Convolution based on the formula of cross-entropy Figure adapted from [32]

consequence less memory for stored parameters and fewer computational operations are needed. Meaningful image features, such as edges, are still detected using a kernel with about 100 pixels, even though the input image contains thousands of pixels.

Parameter Sharing This term refers to the use of the same parameters in more than one function of the model. In *Feedforward Networks* the entries of the weight matrix $\boldsymbol{W}$ of the affine part of a unit are used once, when computing the output of the layer. In a CNN each of the kernel's members are used at several positions of the input. This reduces requirements for parameter storage even further.

Equivariant Representations A function is equivariant, if the input changes and the output changes in the same way. Formally, a function $f(x)$ is equivariant to a function g if

$$f(g(x)) = g(f(x)). \tag{4.21}$$

In case of convolution equivariance to translation is given. Let $\boldsymbol{I}$ be a function giving image brightness in integer coordinates and g a function that shifts the input, such that $\boldsymbol{I}(x, y)' = g(\boldsymbol{I}(x, y)) = \boldsymbol{I}(x - 1, y)$. Then, applying the shift to $\boldsymbol{I}$ and the convolution to $\boldsymbol{I}'$ afterwards is the same as applying the convolution first and then g.
This property is a consequence of parameter sharing and accounts for features that are not bound to a certain location but are found anywhere in the image.

In the context of CNNs convolution refers to an operation, which consists of more then one convolution operations applied in parallel with different kernels to extract different features. Additionally, the input consists not only of a 2-dimensional grid but of more channels. A color image for example has 3 channels, red, green and blue. Those multichannnel operations are calculated as follows. Let $\boldsymbol{K}$ be a 4D tensor with the entry $K_{i,j,k,l}$ giving the connection strength between a unit in channel i of the output and a unit in channel j of the input, with an of set of k rows and l columns between output and input unit. Further, there is the input $\boldsymbol{V}$ with input values V_i, j, k, within channel i at row j and column k and the output $\boldsymbol{Z}$ with the same format as $\boldsymbol{V}$. The entries of $\boldsymbol{Z}$ are then produced by the formula

$$Z_{i,j,k} = \sum_{l,m,n} V_{l,j+m-1,k+n-1} K_{i,l,m,n} \tag{4.22}$$

It is also possible to define a down sampling convolution, if only every s pixels in each direction are sampled in the output

$$Z_{i,j,k} = \sum_{l,m,n} [V_{l,(j-1)s+m,(k-1)s+n} K_{i,l,m,n}]. \tag{4.23}$$

s is also called *stride.*
Even without using strides in the convolving operation, the width of the representation of the input shrinks by one pixel less than the kernel width each layer. *Zero padding* the input of an (intermediate) layer helps to avoid this, by offering the opportunity to control the kernel output size of a layer independently. There are different possibilities to use *zero padding*, for example adding just so many zeros to keep the output's size equal to the size of the input or adding zeros so that each pixel is visited k times in each direction, where k is the kernel width.
Additional to standard convolution (4.22) and strided convolution (4.23) other variants are possible, for example unshared convolution, which is useful for features restricted to a specific region. More details on that are found in [32]. Further, since convolution also has a matrix representation, the transpose of this matrix allows to get a backward pass from the output of the original input [25]. This operation is called *transpose convolution* or *fractionally strided convolution* [25]. A *fractionally strided convolution* corresponds to a convolution operation with *zero padding* [25].

The convolutional operation alone does not built a CNN. A layer consists only in the first stage of convolutions performed in parallel, producing a set of linear activation functions. Each of them runs through a non-linear function, like the *rectified linear activations*, afterwards. This step is referred to as *detector stage.* This step is followed by a so-called *pooling function*, which replaces the output at certain locations within the net with a summary statistic of the outputs in a rectangular neighborhood. The max poling returns the maximum, for example. Pooling is helpful to make representations invariant to small translations in the input.

4.2.3 Basics for training Deep Networks

After having encountered the design of *Feed Forward Neural Networks* and *Convolutional Neural Networks* this section is concerned with the training of *Deep Networks*. Based on a cost function, like that in (4.16), the parameters θ of the neural network f are optimized to approximate a function f^*, which fulfills $\boldsymbol{y} = f^*(\boldsymbol{x}, \theta)$ for an input $\boldsymbol{x}$ and corresponding desired output $\boldsymbol{y}$. One way to do this is gradient based optimization using training data, where for each input $\boldsymbol{x}$ the output $\boldsymbol{y}$ is already defined. In the following *Stochastic Gradient Descent* [32] is described as well as *Backpropagation* [59], which is a procedure to obtain the gradients needed for *Stochastic Gradient Descent* [32].

Stochastic Gradient Descent (SGD) This procedure optimizes the parameters $\boldsymbol{\theta}$ of a model based on a cost function of the form

$$J(\boldsymbol{\theta}) = \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim\hat{p}_{data}} L(f(\boldsymbol{x}, \boldsymbol{\theta}), y) = \frac{1}{n}\sum_{i=1}^{n} L(f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta}), \boldsymbol{y}^{(i)}), \tag{4.24}$$

where $\hat{p}_{data}$ is the empirical distribution of the data, given by the training set, L the per-example loss, $\boldsymbol{x}$ the input, $\boldsymbol{y}$ the output and n the size of the training set. Optimization is done by using its gradient

$$\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \nabla_{\boldsymbol{\theta}} L(f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta}), \boldsymbol{y}^{(i)}), \tag{4.25}$$

however, the gradient is not computed on the whole training set but on a randomly sampled subset of size m, which is called a *minibatch*. The update rule of parameters $\boldsymbol{\theta}$ is given in Algorithm 2.

Algorithm 2 (Minibatch) Stochastic Gradient Descent Update. The algorithm is taken from [32]

```
Require: Learning rate schedule $\epsilon_1, \epsilon_2$ ...
Require: Initial parameter $\boldsymbol{\theta}$
$k \leftarrow 1$
while stopping criterion not met do
    Sample minibatch of $m$ examples of training set $\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(m)}\}$ with corresponding targets $y^{(i)}$.
    Compute gradient estimate:
        $\hat{g} \leftarrow \frac{1}{m}\sum_{i=1}^{m} \nabla_{\boldsymbol{\theta}} L(f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta}), \boldsymbol{y}^{(i)})$
    Apply update $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \epsilon_k \hat{g}$
    $k \leftarrow k + 1$
end while
```

To ensure convergence, a sufficient condition is [32]

$$\sum_{k=1}^{\infty} \epsilon_k = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \epsilon_k^2. \tag{4.26}$$

For cases of high curvature or small but consistent gradient there is a possibility to accelerate the SGD by the method of *momentum*. An additional parameter $\alpha \in [0, 1)$ and a variable $\boldsymbol{v}$ are introduced and the update rule is changed to

$$\begin{aligned} \boldsymbol{v} \leftarrow & \alpha \boldsymbol{v} - \epsilon \nabla_\theta \Big(\frac{1}{m} \sum_{i=1}^{m} L(f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta}), \boldsymbol{y}^{(i)}) \Big), \\ \boldsymbol{\theta} \leftarrow & \boldsymbol{\theta} + \boldsymbol{v} \end{aligned}$$

in case of *Standard Momentum* [4] and to

$$\begin{aligned} \boldsymbol{v} \leftarrow & \alpha \boldsymbol{v} - \epsilon \nabla_\theta \Big(\frac{1}{m} \sum_{i=1}^{m} L(f(\boldsymbol{x}^{(i)}, \boldsymbol{\theta} + \alpha \boldsymbol{v}), \boldsymbol{y}^{(i)}) \Big), \\ \boldsymbol{\theta} \leftarrow & \boldsymbol{\theta} + \boldsymbol{v} \end{aligned}$$

in case of *Nesterov Momentum* [27] as summarized in [32]. $\boldsymbol{v}$ accumulates all the previous gradients. The higher α is relative to the learning rate ϵ, the higher the effect of previous gradients on the current direction. The difference between those two variants is where the gradient is evaluated.

Backpropagation Backpropagation [59] is a method to calculate the gradient of a cost function $\nabla_{\boldsymbol{\theta}} J$. The basic idea behind it is the chain rule of calculus, which is used to compute the derivatives of functions formed as a composition of other functions, whose derivative have already been calculated. For $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{y} \in \mathbb{R}^n$, $z \in \mathbb{R}$ and the mappings $g : \mathbb{R}^m \to \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ the chain rule is given by

$$\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial z}{\partial y_i} \frac{\partial y_j}{\partial x_i}, \tag{4.27}$$

which is equivalent to

$$\nabla_{\boldsymbol{x}} z = \Big(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} \Big)^T \nabla_{\boldsymbol{y}} z, \tag{4.28}$$

with the $n \times m$ Jacobian of g, $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$. This is extensible to tensors of arbitrary size. If $\mathbf{X}$ ia a tensor, the gradient of a value z with respect to $\mathbf{X}$ is denoted by $\nabla_{\mathbf{X}} z$. Let i be a variable, which presents the complete tuple of indices of $\mathbf{X}$, then for all possible index tuples i, $(\nabla_{\mathbf{X}} z)_i$ gives $\frac{\partial z}{\partial \mathbf{X}_i}$. Therefore the chain rule applied to tensors is

$$\nabla_{\mathbf{X}} z = \sum_j (\nabla_{\mathbf{X}} \mathbf{Y}_j) \frac{\partial z}{\partial \mathbf{Y}_j}, \tag{4.29}$$

if $\mathbf{Y} = g(\mathbf{X})$ and $z = f(\mathbf{Y})$.

In the following a *Deep Network* is imagined as a computational graph, where each node indicates a variable. Further the graph consists of operations, which are functions of one or more variables. To give the idea of *Backpropagation* it is assumed for simplicity that an operation returns only a single output variable and each variable is scalar valued. In the graph a directed edge is drawn from variable x to y, if y is computed by an operation applied to x.

Let $\mathcal{G}$ be a graph, which describes the computation of a single scalar $u^{(n)}$, with the input nodes $u^{(1)}$ to $u^{(n_i)}$. For optimizing the performance of $\mathcal{G}$, the scalar's gradient with respect to the input nodes is here the desired quantity or in other words $\frac{\partial u^{(n)}}{\partial u^{(i)}} \forall i \in \{1, 2, ...n_i\}$ have to be computed. If $\mathcal{G}$ describes the forward propagation, in which each node $u^{(i)}$ is computed from the set $\mathbb{A}^{(i)}$, containing all parents of $u^{(i)}$

$$u^{(i)} = f^{(i)}(\mathcal{A}^{()}), \tag{4.30}$$

then let $\mathcal{B}$ be the graph associated with backpropagating the gradient form the output of $\mathcal{G}$ to the input. $\mathcal{B}$ is computed in the reverse order of G and each of its nodes computes $\frac{\partial u^{(n)}}{\partial u^{(i)}}$. If there is an edge from $u^{(j)}$ to $u^{(i)}$ in $\mathcal{G}$, then there is one edge from $u^{(i)}$ to $u^{(j)}$ in $\mathcal{B}$. The computation of the derivative is based on the chain rule

$$\frac{\partial u^{(n)}}{\partial u^{(j)}} = \sum_{i:j \in Pa(u^{(i)})} \frac{\partial u^{(n)}}{\partial u^{(i)}} \frac{\partial u^{(i)}}{\partial u^{(j)}} \tag{4.31}$$

Algorithm 3 describes the backpropagation procedure for graph $\mathcal{G}$. In the overall expression of the gradient, subexpressions would be calculated repeatedly if they are not stored. Therefore the algorithm uses a structure `grad_table` to reduce the needed memory.

Algorithm 3 Backprobagation applied to a graph $\mathcal{G}$ that maps n_i input nodes to a scalar output. Algorithm taken from [32]

```
Run forward propagation to obtain the activations of the network
Initialize grad_table, a data structure that stores the derivatives that have been computed. The entry grad_table[$u^{(i)}$] stores the value of $\frac{\partial u^{(n)}}{\partial u^{(i)}}$
grad_table[$u^{(n)}$] $\leftarrow 1$
for $j = n - 1$ down to 1 do
    grad_table[$u^{(i)}$] $\leftarrow \sum_{i:j \in Pa(u^{(i)})}$grad_table[$u^{(i)}$]$\frac{\partial u^{(i)}}{\partial u^{(j)}}$
end for
return {grad_table[$u^{(i)}$]$| i = 1, ...n_i$}
```

This part on *Backprobagation* is designed to give the basic idea of the algorithm. Much more detail can be found in [32].

4.3 Deep Convolutional Generative Adversarial Nets

After an introduction in *Deep Networks*, the building block of GANs, this section is devoted to the application of GANs to image data. When GANs are introduced in Goodfellow et al.'s paper [70], the authors already experiment with GANs, whose *Generator* and *Discriminator* are represented by *Convolutional Neural Networks*. This model framework is improved by Radford et al. [60], by finding a family of architectures, which is stable to train on a range of datasets. This specialized GAN architecture is called *Deep Convolutional Generative Adversarial Networks* (DCGANs). Architecture guidelines given in [60] are

- Replace any pooling layer with strided convolutions in the *Discriminator* and fractionally strided convolutions in the *Generator*
- For deeper architectures remove fully connected hidden layers
- Use *Batch Normalization* for all layers except the discriminator input layer and generator output layer
  Batch Normalization [15] is a way to stabilize learning by normalizing the input to each component of a layer. It is described in more detail in the next section
- Use Rectified Linear Units activation in the *Generator* for all layers except the output which uses `tanh`
- Use Leaky Rectified Linear Units activation in the discriminator for all layers

Further architecture considerations presented in their paper include that in the *Discriminator* the output of the last convolutional layer is flattend and put into a single sigmoid-function. Besides the input noise $\boldsymbol{z}$ of the *Generator* is first fed though a linear part and then reshaped into a 4-dimensional tensor, to serve as start for the convolutional up-sampling. In Figure 4.3 bellow, there is an example of the *Generator*'s architecture.

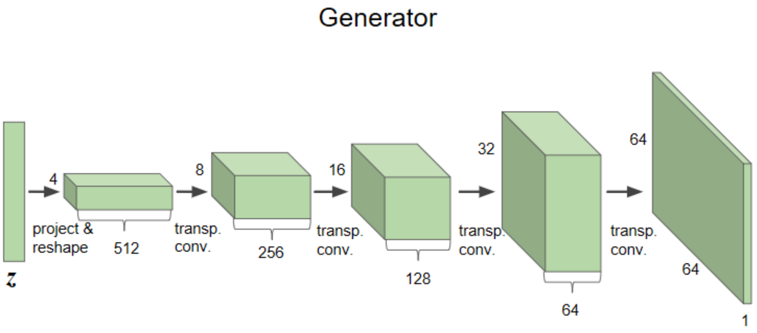

Figure 4.3: Example of Generator's architecture, Figure adapted from [60]

4.3.1 GANs for Image Inpainting

The DCGAN-architecture is used by Yeh *et al.* [48] for image inpainting. The *Generator* G and the *Discriminator* D are trained on uncorrupted data. The hypothesis is that a *Generator* is only able to represent the learned distribution p_{data} but not images not belonging to it. Therefore, to inpaint an image with missing regions, it is searched for an encoding $\hat{\boldsymbol{z}}$ that is closest to the corrupted image. As soon as $\hat{\boldsymbol{z}}$ is found, the recovered image with filled regions is obtained by feeding it in the trained *Generator.* $\hat{\boldsymbol{z}}$ is obtained by

$$\hat{\boldsymbol{z}} = \arg\min_{z}\{\mathcal{L}_c(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{M}) + \mathcal{L}_p(\boldsymbol{z})\}, \tag{4.32}$$

where $\mathcal{L}_c$ is the context loss, which constraints the generated image given the hole mask $\boldsymbol{M}$, indicating missing parts and the image being filed $\boldsymbol{y}$. $\mathcal{L}_p$ forces the generated image to look realistic. To find the correct encoding more attention is paid to pixels close to holes than to pixels, which are far away from them. This is formalized by a weighting term $\boldsymbol{W}$. The contextual loss is the given by

$$\mathcal{L}_c(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{M}) = ||\boldsymbol{W} \odot (G(\boldsymbol{z}) - \boldsymbol{y})||_1, \tag{4.33}$$

where $\odot$ denotes the element-wise multiplication. $\mathcal{L}_p$ is defined using the output of the *Discriminator* D,

$$\mathcal{L}_p = \lambda \log(1 - D(G\boldsymbol{z})) \tag{4.34}$$

and λ is used to balance between the two losses. The inpainted image $\hat{\boldsymbol{x}}$ is then obtained from $G(\boldsymbol{z})$, using its gradients to shift image colors to match the colors of $\boldsymbol{y}$.

4.3.2 GANs for Anomaly Detection

The idea to inpaint images using GANs is adapted to detect anomlies in 3-dimesional image volumes of the retina by Schlegl et al. [28]. For this purpose G and D are trained on image patches of size 64×64 extracted from healthy subjects. After training is completed, image patches not seen during training are mapped to the space of random noise inputs and the *Generator* creates an artificial healthy version of the input image. The loss function on which the mapping is based is then used as anomaly score. Since the *Generator* is only trained on images showing healthy tissue, the hypothesis is that the mapping is more accurate for images not showing lesions that for those which do. Therefore the loss function has higher values for patches showing anomalies than those which do not.

Schlegl et al. [28] define the loss function differently compared to the inpainting approach. It is composed of a *residual loss* $\mathcal{L}_R$ which corresponds to the *contextual loss* L_c and a *discrimination loss* $\mathcal{L}_D$ which corresponds to $\mathcal{L}_p$. For the *residual loss*, the weighting term $\boldsymbol{W}$ is not included but it also forces the generated image to look similar to the encoded image,

$$\mathcal{L}_R(\hat{\boldsymbol{z}}) = \sum |\boldsymbol{x} - G(\hat{\boldsymbol{z}})|. \tag{4.35}$$

The *discrimination loss* is not based on the output of the *Discriminator*, like $\mathcal{L}_p$ but on the output of intermediate layer f of the *Discriminator*,

$$\mathcal{L}_D(\hat{\boldsymbol{z}}) = \sum |f(\boldsymbol{x}) - f(G(\hat{\boldsymbol{z}}))|. \tag{4.36}$$

This is inspired by the feature matching technique [67] and forces the generated image to match the features of images showing healthy tissue.

4.4 Wasserstein GAN and further improvements for GAN-Trainig

Whereas the previous section is concerned with the architecture of DCGANs and how they are used to detect anomalies in 2D patches, this section is devoted to the description of two ways to stabilize the training of GANs in general. The first one is *Batch Normalization* [15], a procedure which is recommended to be used by Radford et al.[60]. The second is a changed training process compared to the one proposed in Godfellow et al. [70], which is called *Wasserstein GAN* (WGAN) [46]. Both, reduce the appearance of mode collapse [46], [60]. This refers to the problem that different inputs of G yielding the same output [70].

4.4.1 Batch Normalization

Batch normalization is a way to standardize the input of any layer of a neuronal network [32] and is introduced in [15]. Let $\boldsymbol{H}$ be a design matrix, where the columns correspond to a minibatch of units of the layer. The inputs for each example are in the rows of $\boldsymbol{H}$. To normalize the design matrix it is substituted by $\boldsymbol{H}'$ [32]

$$\boldsymbol{H}' = \frac{\boldsymbol{H} - \boldsymbol{\mu}}{\boldsymbol{\sigma}}, \tag{4.37}$$

where

$$\boldsymbol{\mu} = \frac{1}{m}\sum_i \boldsymbol{H}_{i;:} \quad \text{and} \quad \boldsymbol{\sigma} = \sqrt{\delta + \frac{1}{m}\sum_i (\boldsymbol{H} - \boldsymbol{\mu})_i^2}, \tag{4.38}$$

where δ is a small positive value such as 10^{-8} to avoid the term under the square root to become 0, since the gradient of $\sqrt{z}$ is not defined at $z = 0$ [32]. A defined gradient is needed for *Backpropagation* [32].

4.4.2 Wasserstein GAN

GANs are used to learn a distribution p_{data} by training the *Generator* G, so that p_g becomes equal to p_{data} [70]. p_g is the distribution of examples that G creates. The training procedure of *Wasserstein GANs* makes use of the so-called *Earth Mover* (EM) distance, to turn p_g into p_{data} [46]. Arjovsky et al. provide theoretical background concerning the EM distance and give an algorithm for training *Wasserstein GANs* in their paper [46]. Essential results for understanding the WGAN training algorithm are summarized here.

The EM distance W between two distributions p_g and p_{data} is defined as

$$W(p_{data}, p_g) = \inf_{\gamma \in \Pi(p_{data}, p_g)} \mathbb{E}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \gamma}[||\boldsymbol{x} - \boldsymbol{y}||], \tag{4.39}$$

where $\Pi(p_{data}, p_g)$ denotes the set of all joint distributions $\gamma(p_{data}, p_g)$ whose marginals are respectively p_{data} and p_g. Another way to define W is

$$W(p_{data}, p_g) = \sup_{||f||_L \leq 1} \mathbb{E}_{\boldsymbol{x} \sim p_{data}}[f(\boldsymbol{x})] - \mathbb{E}_{\boldsymbol{x} \sim p_g}[f(\boldsymbol{x})], \tag{4.40}$$

where the supremum is over all 1-Lipschitz functions $f : X \to \mathbb{R}$. The latter definition is more practicable for the WGAN-algorithm.
In practice, the EM distance $W(p_{data}, p_g)$ is the cost function based on which the GAN is trained. The *Generator* G is optimized using the gradient $\nabla_{\boldsymbol{\theta}_g} W(p_{data}, p_g)$ when $\boldsymbol{\theta}_g$ are parameters of G, which are optimized during training. To compute this gradient the f which is the supremum in (4.38) has to be found. The *Discriminator* D therefore takes its role and is found via optimizing parameters $\boldsymbol{\theta}_d$ so that

$$\max_{\boldsymbol{\theta}_d \in \boldsymbol{\Theta}} \mathbb{E}_{\boldsymbol{x} \sim p_{data}}[D(\boldsymbol{x}; \boldsymbol{\theta_d})] - \mathbb{E}_{\boldsymbol{z} \sim p_z}[D(G(\boldsymbol{z}, \boldsymbol{\theta}_g), \boldsymbol{\theta}_d)] \tag{4.41}$$

is fulfilled. The parameter space $\boldsymbol{\Theta}$ has to be compact to ensure Lipschitz- continuousness. This is enforced by clipping the parameters θ_d after every update. The full algorithm is given in Algorithm 4.

Beside the cost function, the WGAN algorithm relies on RMSProp as optimization algorithm in contrast to traditional GANs [70], where Adam [38] is used. Further, the *Discriminator* does not fulfill its original function to discriminate between samples from the data distribution and generated examples by assigning a label. But it still helps the *Generator* to be trained by providing an approximation of the EM distance $W(p_{data}, p_g)$.

Algorithm 4 Training Generative Adversarial Networks using using the EM distance as cost function, The algorithm is adapted from [46]

```
for number of training iterations do
    for $k$ steps do
        Sample minibatch of $m$ noise samples $\{\boldsymbol{z}^{(1)}, ..., \boldsymbol{z}^{(m)}\}$ from noise prior $p_z$.
        Sample minibatch of $m$ examples $\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(m)}\}$ from $p_{data}$.
        Update the Discriminator based on its stochastic gradient:
            $$\nabla_{\boldsymbol{\theta}_d} \frac{1}{m} \sum_{i=1}^{m} [D(\boldsymbol{x}^{(i)}) - D(G(\boldsymbol{z}^{(i)}))].$$
        clip($\boldsymbol{\theta_d}; -c, c$)
    end for
    Sample minibatch $m$ noise samples $\{\boldsymbol{z}^{(1)}, ..., \boldsymbol{z}^{(m)}\}$ from noise prior $p_z$
    Update the Generator based on its stochastic gradient
        $$-\nabla_{\boldsymbol{\theta}_g} \frac{1}{m} \sum_{i=1}^{m} D(G(\boldsymbol{z}^{(i)})).$$
end for
```

4.5 Discussion

In this chapter *Generative Adversarial Networks* and its building blocks *Deep Networks* have been described. GANs are composed of two adversarial components a *Generator* and *Discriminator*, which in its traditional form, play a min-max game against each other to optimize their abilities. Both are *Deep Networks*, one designed to learn a data distribution and the other to support the learning procedure. Several forms of Deep Networks exist, including *Feedforward Neural Networks* and *Convolutional Neural Networks*. The first ones are the simplest form of a neural network, the latter are a specialized form, which are applied in connection with images and therefore of special interest for this thesis. For both cases various design possibilities exist, described in 4.2.1 and 4.2.2. *Deep Networks* and GANs are trained on gradient based optimization algorithms, such as *Stochastic Gradient descent* and its variants. The gradients needed for these algorithms are calculated by a cost function using *Backpropagation*.

To apply GANs on image data, guidelines for architecture design of *Deep Convolutional Generative Adversarial Networks* have been given and its application to anomaly detection has been described. Finally, two ways to stabilize the training of GANs have been discussed, *Batch Normalization* and *Wasserstein GAN*. The latter changes the training procedure of GANs by introducing the *earth mover* distance as new cost-function.

CHAPTER 5

Specific Related Work

In this chapter an overview of state-of the art approaches of the corner stones of this thesis are given. Whereas the previous chapters give an introduction to the methodology used in this thesis, this chapter focuses on solutions provided on registration of breast images, the segmentation of the whole breast and the detection of lesions using different approaches. Breast registration and anomaly detection make up a larger part of this thesis than segmentation. Therefore the section about the latter topic is kept short and gives only a few examples.

5.1 Breast Registration

Papers dealing with the registration of breast images involving DCE-MR images discuss three different fields. One is inter-modal registration, where MRI, Computational Tomography (CT), mammography or Positron Emission Tomography (PET) images are registered [44], [62]. The second group of papers is concerned with intra-modal registration of DCE-MR images belonging to the same scanning series [62], [17]. The aim is to register images acquired after the injection of a contrast agent to the pre-contrast image. The last class of papers deals with the intra-modal registration for treatment planning after surgical removal of a tumor [53]. In this case DCE-MR images acquired before and after the surgery are registered.

In his review article in 2006, Gou *et al.* [62] summarizes the current status of breast registration at that time, including intra-modal and inter-modal registration. For registration of 2D X-ray mammograms to 3D DCE-MRI images, Ruiter *et al.* [65] apply a finite element model of the deformable behavior of the breast. More recent examples for the application of biomechanical models in breast registration also exist, with an overview given in [52]. A registration system for PET-CT and MRI images has also been prosed [44].
Intra-modal registration approaches to register DCE-MR images include feature-based

and intensity-based registration approaches, with an overview in [62]. Feature based methods rely on the selection and matching of control points, whereas intensity based approaches are built on different assumptions regarding image intensity and partly include only rigid transformation [62], [17]. In the paper of Rueckert *et al.* [16] a global affine transformation followed by a local free form registration model based on B-splines is used combined with *Mutual Information.* Kuczynski *et al.* [54] also use B-splines but in combination with the *Mean Squared Distance* as matching criterion and an optimizer suitable for bound problems. Kim *et al.* [17] choose a different approach by registering post contrast images group-wisely to the group-mean image and registering the latter to the pre-contrast image.

Wodzinski *et al.* perform registration of images across points in time using variants of the *Demons* Algorithm and comparing them to free form deformation using B-splines [53]. They conclude that *Symmetric Diffeomorphic Demons* are more suitable for small deformations and B-splines perform better if deformations are large. However, they also see space for specialized algorithms.

In this thesis *Diffeomorphic Demons* are used to align 2 images, which belong to the same patient but have been acquired at different points in time. Nonrigid registration is necessary to compensate the change that breast tissue undergoes over time [24].

5.2 Breast Segmentation

Segmentation of the whole breast is a pre -processing step for anomaly detection to reduce computational effort [14]. In [14] a brief overview of different segmentation approaches is given. They themselves propose an algorithm consisting of several steps. These steps include the use of Otsu's Threshold for a primal mask followed by the definition of key points for the characterization of anatomical geometry. Based on these anatomical key points and 2 dimensional clustering, probabilities of voxels belonging to breast tissue are calculated. From this probability map a binary mask is calculated using again Utsu's Threshold. The mask creation is done slice-wise.

Giannini *et al* [5] propose to segment the breast by first distinguishing air and body parts present in the image and then identify the breast-pectoral muscle boundary. This muscle is an anatomical lower bound of the breast. The segmentation of the muscle is done based on anatomical considerations and the intensity gradient of muscle tissue and surrounding tissue.

The approach used in this thesis is inspired by the one used in [45]. For each woman a template breast model most similar in form and size is chosen by registering different breast templates to the patient's MR image.

5.3 Anomaly and Lesion Detection

Several approaches to identify lesions in the breast exist. There are methods which are based on a single imaging modality [11], [10], [34] as it is the case for this thesis but also multi-modal approaches are found in the literature [49], [31]. This section is mainly focused on mono-modal approaches, especially those performed on DCE-MR images.

In the literature there are several papers in which lesions are classified based on features. Differences occur in the features and classifiers used. Yao *et al.* [11] use textural analysis of the temporal DCE-MR sequence and velvet transformation to obtain features for pixel-wise classification by a Support Vector Machine (SVM). Levman *et al.* [7] also use a SVM for classification but signal intensity features and Gubern-Mérida *et al.* [10] use signal enhancement and blob features and compare different classifiers. The Random Forest classifier outperformed the other classifiers, including a SVM classifier. Other papers focus on the detection of specific tumor classes such as non-mass lesions [34] or triple negative cancer [41]. Platel *et al.* [57] investigate if an ultra fast DCE-MR imaging protocol allows the same classification performance as an imaging protocol with regular length based on morphological and kinetic features and different classifiers such as kNN, Random Forests and two variants of SVMs.

Approaches including deep learning have also been investigated, including a CNN as feature extractor or as classifier and cross-modal transfer learning. Selvathi *et al.* [6] use a CNN to extract features from mammograms and compare classifiers such as kNN, SVMs and Random Forests. In the paper of Zhang *et al.* [39] CNNs with different architectures are trained to classify lesions in mammograms. A CNN trained on natural images is used to create a feature map for the mammograms. These feature maps are used to train the CNNs for the classification task. Transfer learning is also used by Hadad *et al.* [26]. They train CNN-variants on mammograms and use them to classify lesions in DCE-MR images.

5.4 Discussion

In this chapter scientific work in the fields of breast DCE-MR image registration, breast segmentation, and anomaly detection has been presented. In case of registration, approaches for multi and mono-modal registration have been described. For segmentation 3 examples have been given. Anomaly detection of lesions in the breast has been carried out by the definition of features and classification based on them in the past. CNNs have also served as classier.

To the author's knowledge a GAN has not been used to detect lesions in the breast so far.

CHAPTER 6

Methodology

This chapter is devoted to the description of the methodology of this thesis. The aim is the detection of lesions in the female breast using a GAN. The GAN does not serve directly as classifier but learns the distribution of image patches of healthy tissue. This knowledge is then used to recognize anomalous tissue as divergent from this distribution resulting in an anomaly score.
For each patient, DCE-MR images belonging to more than one point in time are available. Image registration is performed to bring images belonging to the same patient into the same frame. Lesion detection is carried out on difference images calculated between two time points. Therefore, spatial correspondence is necessary.
Beside the pure classification task, the approach described in this thesis is also used for the prediction of lesions. Here we evaluate, if the method is able to recognize lesions before a radiologist has been able to do it.

In the following the steps from raw data to anomaly detection based on GANs are described. The first section offers a description of the raw data, which is used for testing and the evaluation of the presented approach. This is then followed by a description of the first step of the processing pipeline, namely the registration of images. Afterwards the third section briefly describes how the breast tissue is distinguished from background and other anatomical structure not belonging to the breast. The fourth section then deals with the training of a GAN and how it is used to identify anomalies. Finally, the last section discusses the presented approach.

6.1 Data and Preprocessing

DCE-MRI data of 20 high risk patients for breast cancer has been received from the General hospital of Vienna (AKH). As already mentioned, for each patient several points in time at which data have been acquired, exist. The overall range of image acquisition varies from 2 years to around 10 years. In this time period MRI scanners and acquisition

protocols have been changed, leading to 3 modalities with which it is dealt in this thesis. Additionally, not only the amount of acquisition time points varies between patients but also the time periods between two acquisition dates. Some patients came annually to a control of their cancer status, others led 5 years pass between two acquisitions.
In the following the three modalities are described. In Figure 6.1 examples for each, belonging to the same patient are shown.

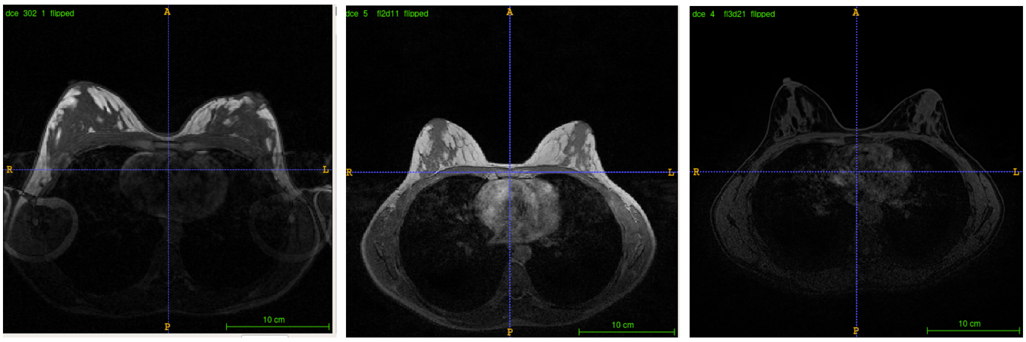

Figure 6.1: Examples of the 3 modalities

Images of the first modality (Figure 6.1, left) are acquired between 2004 and 2007 by a Phillips scanner. Images are of size $265 \times 265 \times 30$ and the voxel spacing varies in z-direction between 4.0 and 4.3 and in $x-$ and $y-$ direction between 1.25 and 1.5. An acquisition series consists of one pre-contrast image and 7 post-contrast images. The time period between pre-contrast image and first post contrast image is 2 minutes. Then, always after one minute the next post-contrast image is acquired.
Between 2007 and 2014 Modality 2 (Figure 6.1, middle) was in use. Images are of size $384 \times 384 \times 48$ and the voxel- spacing is 0.9115 in x- and y-direction and 3.3 in z-direction. One acquisition time point is an exception with a size of $384 \times 384 \times 52$. Images were acquired by a Siemens scanner and an acquisition series consists of one pre-contrast image and 5 post-contrast images, with 2 minutes between each image acquired.
The newest and third modality has been in use since 2014 and produces images of size $512 \times 512 \times 80$ by a Siemens scanner. The voxel-spacing is 0.7031 in x- and y- direction and 2 in z-direction. An acquisition consists beside the usual pre-contrast image of 3 post-contrast images. The time span between the acquisition of pre-contrast and first post-contrast images is 2 minutes. Then the time span is 1.5 minutes. Compared to the other two modalities the signal of fatty tissue is suppressed. Table 6.1 summarizes the most important properties of the three modalities.

Table 6.1: Properties of the 3 modalities

	Modality 1	Modality 2	Modality 3
years in use:	2004-2007	2007-2014	2014-2017
size:	$265 \times 265 \times 30$	$384 \times 384 \times 48$	$512 \times 512 \times 80$
No. post-contrast images:	7	4	3
time betw. post-contrasts:	1 min	2 min	1.5 min
time pre- and 1^{st} post contrast:	2 min	2 min	2 min

Preprocessing The patients' data has been pseudonymized and received in DICOM format [18]. The first preprocessing step was therefore the conversion of the data to nifti format (https://nifti.nimh.nih.gov/) using dcm2niix [1]. For further processing and the anomaly detection, the data needed for each patient are the pre-contrast images of each acquisition series and the difference images of post-contrast images and pre-contrast image. Those difference images are already part of the data set except for two time points of two different patients. In the latter case, the difference images have been calculated by simply subtracting the pre-contrast image from the post-contrast images using ImageMath, which is part of the ANTs registration suite [36]. Finally, prior to image registration a bias field correction algorithm has been applied [37], whose implementation is also part of the ANTs registration suite [36]. The registration pipeline is described in the next section.

6.2 Image Registration

After bias field correction, the image registration is the next processing step. The data processed for each patient in this section are the pre-contrast images $\boldsymbol{I}_{t,m}^{pre}$ of each acquisition series t and the difference images $\boldsymbol{I}_{t,m,pt}^{difference}$ of post-contrast images and pre-contrast image. The parameter t identifies the acquired series by its chronological order starting at 1 and ending with the number of acquisition series available for the actual patient, whereas $m = 1, 2, 3$ indicates the modality with which the series has been acquired. The parameter *pt* is short for **p**ost-contrast **t**ime and yields the information how many minutes after contrast agent injection the corresponding post-contrast image has been acquired. For example, if $pt = 2min$ the post-contrast image has been acquired 2 minutes after the injection.
The aim is to reach spatial correspondence between images, which belong to the same patient. Therefore the registration framework is described only for one patient.

Basic Registration Scheme The target frame is the one of the chronologically last acquisition series. If t_n time points are available for the patient, then the goal is to bring all images into the frame of $\boldsymbol{I}_{t_n,m}^{pre}$. Pre-contrast images $\boldsymbol{I}_{t,m}^{pre}$ and images $\boldsymbol{I}_{t,m,pt}^{difference}$, which belong to the same acquisition series t share a common frame. Therefore, all images belonging to series t_i are brought into the frame of $\boldsymbol{I}_{t_n,m}^{pre}$ by registering $\boldsymbol{I}_{t_i,m}^{pre}$

to $\boldsymbol{I}^{pre}_{t_n,m}$ and applying the obtained transformation to all images belonging to series t_i. Figure 6.2 illustrates the registration of an acquisition series to the target frame.

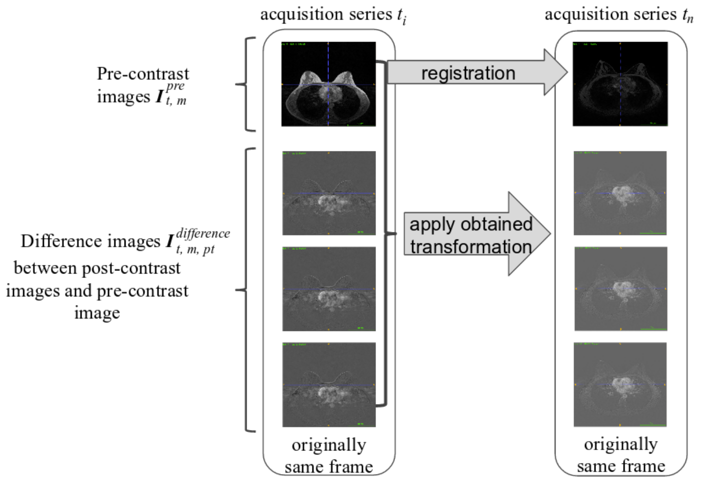
acquisition series t_i
acquisition series t_n
Pre-contrast images $\boldsymbol{I}^{pre}_{t,\, m}$
registration
Difference images $\boldsymbol{I}^{difference}_{t,\, m,\, pt}$ between post-contrast images and pre-contrast image
apply obtained transformation
originally same frame
originally same frame

Figure 6.2: Schematic overview of the basic registration scheme for a patient. Images acquired at time point t_i are brought into the frame of the chronologically last time point t_n.

After registration the pre-contrast images $\boldsymbol{I}^{pre}_{t,m}$ of each acquisition series t as well as the difference images $\boldsymbol{I}^{difference}_{t,m,pt}$ of post-contrast images and pre-contrast image are in the same frame, which is the aim of this section. However, it has not been discussed how the transformations are calculated. There are differences between the three modalities which are described in the following. Furthermore, a closer look is taken at the implementation of a transformation.

The modality-dependent Registration Pipeline For a better understanding of the modality-dependent registration pipeline, it is explained using an illustrative example.

For a patient, of whom images were acquired once every year, for example from 2005 to 2016, two time points were acquired with Modality 1 (in 2005 and 2006), 7 points in time with Modality 2 (2007-2013) and 3 points in time with Modality 3 (2013-2016). Calculations of transformations are always done for the pre-contrast image of an acquisition series. Therefore the goal is to bring all images of the patient into the frame of the pre-contrast image of the last time point 2016. This image is denoted $\boldsymbol{I}^{pre}_{12,3}$ where 3 is an indication for Modality 3.
Another image, which serves as a reference image in the registration pipeline is the pre-contrast image of the last time point acquired with Modality 2. For the example patient, this is the pre-contrast image acquired in 2013 and is denoted as $\boldsymbol{I}^{pre}_{t=9,2}$. For all acquisition series, which were produced before $\boldsymbol{I}^{pre}_{t=9,2}$, transformations are calculated to register the images belonging to those series to $\boldsymbol{I}^{pre}_{t=9,2}$. Transformations are calculated

on the pre-contrast images and then applied to the difference images between post- and pre-contrast images $\boldsymbol{I}^{difference}_{t,m,pt}$, since images belonging to the same acquisition series are originally in the same image frame.
If the data for the patient had only been acquired by Modality 1 and 2 and not Modality 3, the registration of this patient would be finished because then all images would be in the frame of the pre-contrast image of the last time point. But, for the example presented here, images have also been acquired with Modality 3. Therefore, a transformation is calculated to register $\boldsymbol{I}^{pre}_{t=9,2}$ to $\boldsymbol{I}^{pre}_{t=12,3}$ and is applied to images belonging to the same point in time as $\boldsymbol{I}^{pre}_{t=9,2}$ and all images, which belong to prior time points after they have been brought into the frame of $\boldsymbol{I}^{pre}_{t=9,2}$. Image series, which were acquired using Modality 3 are registered directly to $\boldsymbol{I}^{pre}_{t=12,3}$, by calculating the transformation based on the pre-contrast images and applying it to all images belonging to the corresponding acquisition series.
In Figure 6.3 and illustration of the registration procedure for one patient is given.

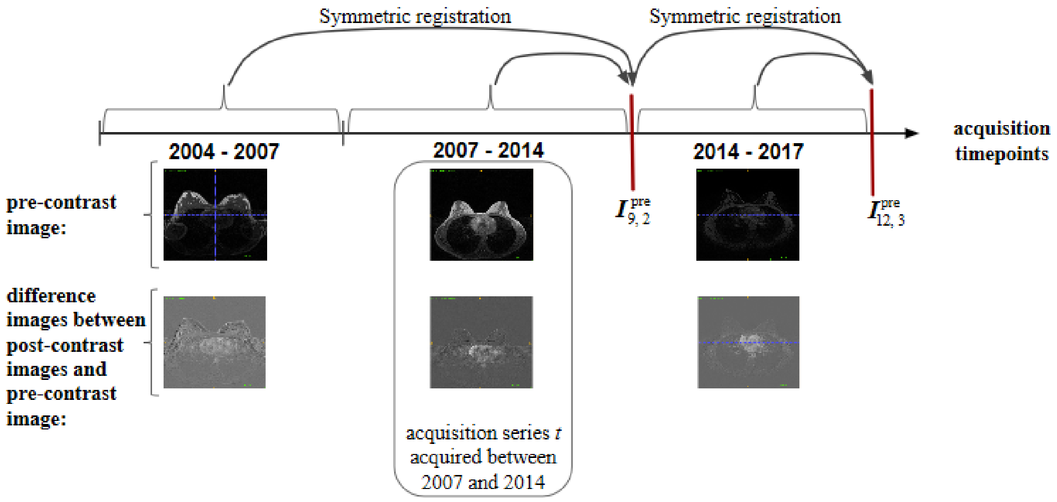

Figure 6.3: Overview of the registration pipeline for a patient, of whom images where acquired b all 3 modalities. Without Modality 3, images are brought into the frame of $\boldsymbol{I}^{pre}_{t=9,2}$.

Implementation The transformation calculated to register one image to another one consists of an affine transformation performed by the ANTs registration software [36] followed by the application of the Diffeomorphic Demons algorithm Ezys [55]. In cases where the registration algorithm gets stuck in a local minimum yielding a nonsensical transformation, a initial alignment of the center of mass has been incorporated later on. Calculations are done using MATLAB 2011a. Ezys performs automatically an affine transformation before the non-rigid registration, however the affine transformation calculated using Ezys alone gives a too weak initialization for the non-rigid transformation

and results get partly unusable. The additional affine transformation performed by ANTs [36] helps to obtain an initialization good enough for the non-rigid transformation part. The choice of parameters are discussed in the next chapter.
In Figure 6.4 examples of registered images are given. The figure shows the result of a patient, where Modality 1 and Modality 2 images have been registered to the pre-contrast image of the last series acquired by Modality 3.

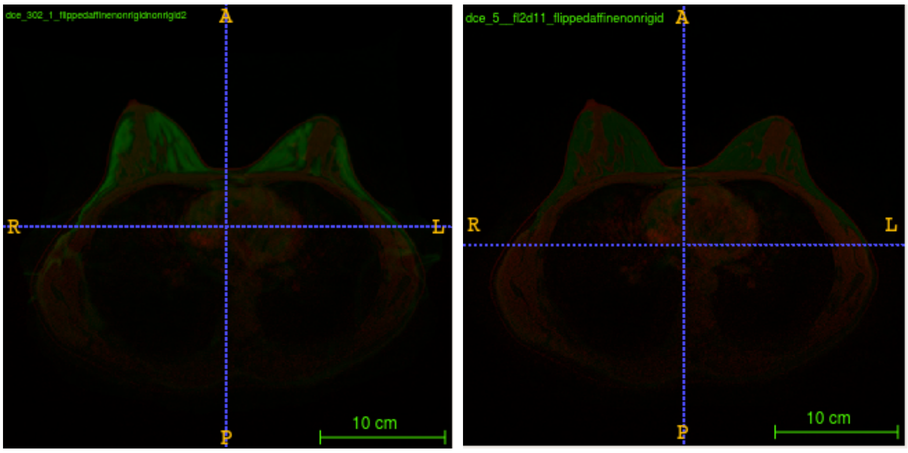
dce_302_1_flippedaffinenonrigidnonrigid2
A
R
L
P
10 cm
dce_5__fl2d11_flippedaffinenonrigid
A
R
L
P
10 cm

Figure 6.4: Registration results for inter-modal registration with Modality 3 in the red and Modality 1 and 2 in the green channel, right: Modality 1 on Modality 3, left: Modality 2 to Modality 3

Ants and Ezys as well use a coarse to fine scheme as described in Section 3.4. and the matching criteria is *Mutual Information* in case of the affine ANTs registration and *Normalized Mutual Information* for the transformations calculated by Ezys. Furthermore, for the transformation calculated by Ezys, symmetry is incorporated by averaging direct $\boldsymbol{I}^{pre}_{t_i,m_l} \rightarrow \boldsymbol{I}^{pre}_{t_j,m_k}$ and indirect registration $\boldsymbol{I}^{pre}_{t_j,m_k} \rightarrow \boldsymbol{I}^{pre}_{t_i,m_l}$, if matching is performed for two images not belonging to the same modality, so $t_i \neq t_j$ and $m_l \neq m_k$ (compare Figure 6.3). Then the transformation $\mathcal{T}_{\boldsymbol{I}^{pre}_{t_i,m_j} \rightarrow \boldsymbol{I}^{pre}_{t_j,m_k}}$ is given as [55]

$$\mathcal{T}_{\boldsymbol{I}^{pre}_{t_i,m_j} \rightarrow \boldsymbol{I}^{pre}_{t_j,m_k}} = \frac{\mathcal{T}_1 + \mathcal{T}_2}{2}, \tag{6.1}$$

where direct registration yields in $\mathcal{T}_1$ and indirect in $\mathcal{T}_2$. This symmetry option is implemented in Ezys counteract bias introduced by registration of an interpolated source image.

6.3 Whole Breast Segmentation

Segmentation of the whole breast is performed to be able to distinguish breast tissue from background and other anatomical structure. Since all pre-contrast images $\boldsymbol{I}^{pre}_{t,m}$ and difference images $\boldsymbol{I}_{t,m,pt}^{difference}$ belonging to the same patient are brought into the same image frame through registration, the creation of one breast mask is necessary per patient and applicable to all those images. In the following methods and implementation of breast mask creation for a patient are described. Segmentations are calculated based on the chronologically last pre-contrast image, which has been acquired by Modality 2. It is denoted as $\boldsymbol{I}_{t=last_2,2}^{pre}$.

Breast Segmentation based on Breast Models As mentioned in Chapter 5, the segmentation is performed similar to the approach in [45]. Out of 9 breast models with different forms and sizes the one, which fits the patient's breast form best, is chosen and the breast mask of the chosen model serves as mask for the patient's breast. To automatically choose a breast model for a patient, each of the 9 templates is registered to $\boldsymbol{I}_{t=last_2,2}^{pre}$ or to its its registered version, if existent. The best model is then chosen through the dice coefficient,

$$DSC = \frac{2|\boldsymbol{S}_{binary} \cap \boldsymbol{T}_{binary}|}{|\boldsymbol{S}_{binary}| + |\boldsymbol{T}_{binary}|} \tag{6.2}$$

after thresholding the reference and source image to obtain a binary version of each image. After the template is chosen, the transformation calculated through registration is applied to the corresponding breast mask to obtain a mask for the patient. An example of a breast segmented using the described approach is given in Figure 6.5, left.

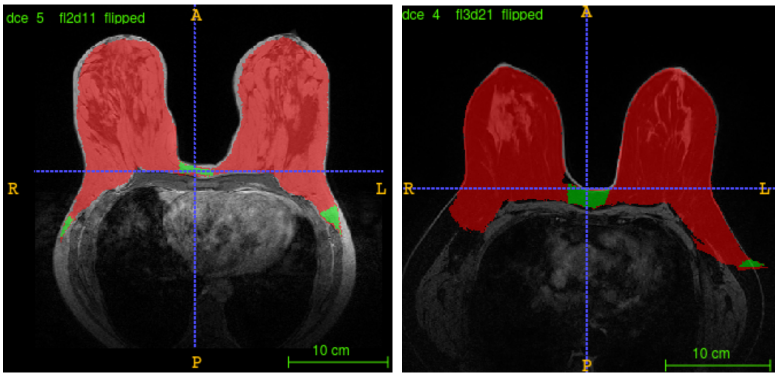

Figure 6.5: Examples of segmented breasts: The segmentation for the breast in the left images have been obtained by using water images as breast templates and $\boldsymbol{I}_{pre,last,3}$ as reference image. For the breast in the left image the segmentation has been obtained by the described approach.

The approach described has been applied to all patients except one, for whom there are not any images available acquired by Modality 2 (or 1). Therefore, for this single patient, the pre-contrast image of the chronologically last acquisition series acquired by Modality

3, $\boldsymbol{I}^{pre}_{last_3,3}$, has been used as reference image and templates have not been fused water-fat images but images showing only the MR signal of water. The result is shown in Figure 6.5, right.

Implementation The breast models and corresponding mask have been available to the author and did not have to be created. One breast model consists of a template image, which is a fused water-fat image, calculated using the decision function

$$\boldsymbol{I}_{WF}(x,y,z) = \frac{1}{1+\exp(\frac{-\boldsymbol{I}_W(x,y,z)-0.25}{\boldsymbol{I}_F(x,y,z)})} \tag{6.3}$$

and a 3D breast mask. In Figure 6.6 an example of a breast template and breast mask is given.

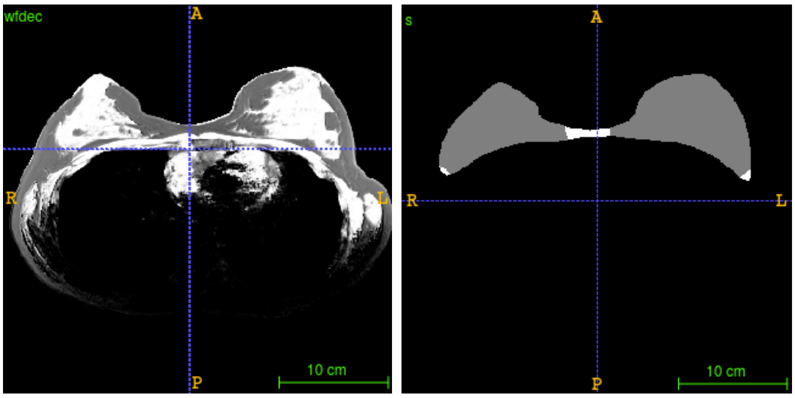

Figure 6.6: Example of breast template (left) and mask (right)

Most parts of the code used have been provided by the author of [66]. Originally, the registration included only an affine transformation using ANTs [36]. This has been extended in this work by the additional calculation of a non-rigid transformation using Ezys [55].

A the end of this section a breast mask for each patient is available and is applicable to the registered versions of all difference images $\boldsymbol{I}^{difference}_{t,m,pt}$ for every acquisition series t. Besides, through registration spatial correspondence between the difference images $\boldsymbol{I}^{difference}_{t,m,pt}$ is archived.

6.4 Anomaly detection

This section describes the use of a *Generative Adversarial Network* for anomaly detection. The whole detection framework consists of training the GAN, down-sampling of new images to latent space and finally the detection of anomalies itself. Each part is discussed in the following from the creation of a training set for the GAN to the definition of an anomaly score, based on which image patches are classified as either showing healthy or anomalous tissue. Figure 6.7 gives an overview of the process.

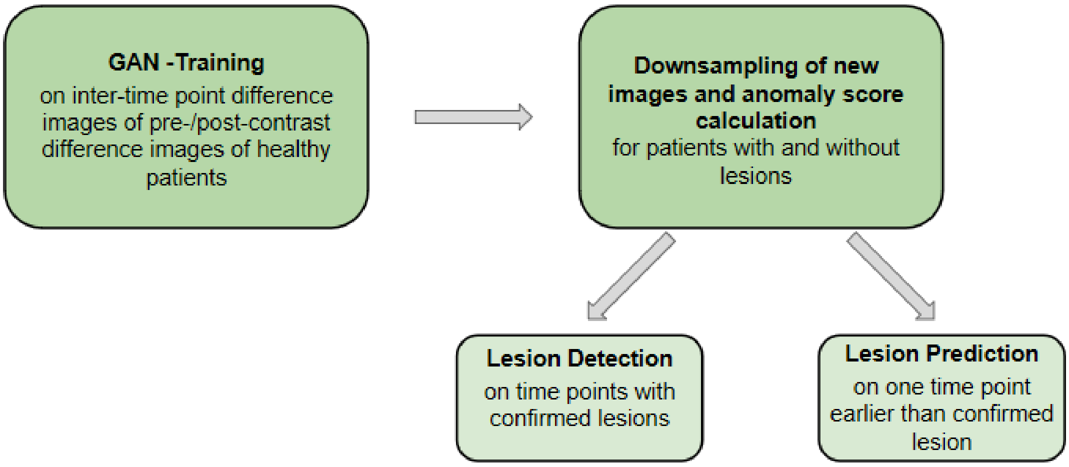
GAN -Training
on inter-time point difference images of pre-/post-contrast difference images of healthy patients
Downsampling of new images and anomaly score calculation
for patients with and without lesions
Lesion Detection
on time points with confirmed lesions
Lesion Prediction
on one time point earlier than confirmed lesion

Figure 6.7: Overview of the anomaly detection and prediction methodology

6.4.1 Creation of a Training Set

The GAN is trained on 2D image patches of size 64 × 64 extracted from patients without lesions. First, for each patient used for the training set, difference images between two acquisition series t_i and t_j are calculated using the difference images $\boldsymbol{I}^{difference}_{t,m,2min}$ calculated between first post-contrast image and pre-contrast image of an acquisition series. The first post-contrast images of an acquisition series are acquired 2 minutes after contrast agent injection, therefore the parameter pt is set to $2min$.
Those difference images of difference images are either calculated between successive time points, t_i and t_{i+1} or with one time point in between, t_i and t_{i+2}. For the rest of this thesis difference images of difference images are denotes as $\boldsymbol{I}^{DiffOfDiff}_{t_i-t_j}$ with $t_i > t_j$, where $t_i - t_j$ holds the information between which time points the difference images are calculated.
After the calculation of inter-time point difference images, the patches are extracted slice-wise from each of the calculated difference images. The slices correspond to sagittal planes through the human body. The center of the 64 × 64 patches is chosen randomly and only patches, which contain 50 or less pixels outside of the patient's breast mask become part of the training set. Before being saved in png-format, the gray values are scaled to the range 0-250. Figure 6.8 shows examples of extracted patches.

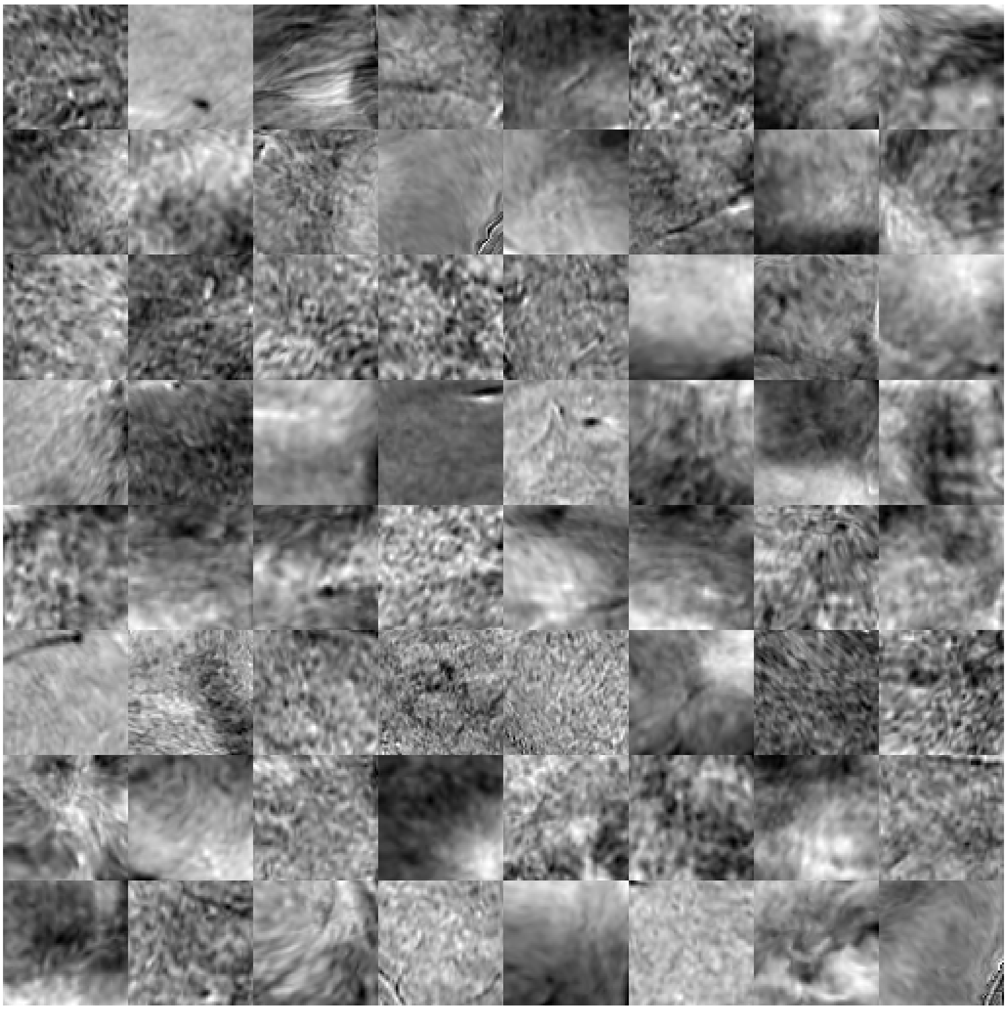

Figure 6.8: Extracted image patches showing healthy breast tissue

6.4.2 GAN-Architecture and Training Procedure

The architecture of the *General Adversarial Network* follows the DCGAN -architecture of [60], which is also described in 4.3. The implementation is done using Python 2.7.13 and TensorFlow [68] and is an adaption from the code provided by [9].

In the *Generator*, the random noise input vector $\boldsymbol{z}$ is fed through a linear part, reshaped to a 3D tensor with dimension $4 \times 4 \times 512$ and then put through a *Rectified Linear Units* layer to serve as start for a convolutional stack. This stack is then fed through four layers of transposed convolution, each followed by a *Rectified Linear Units* layer except the fourth. The tensor dimensions thereby change from $4 \times 4 \times 512$ to $64 \times 64 \times 1$ by a increase of the first 2 tensor dimensions and a decrease of the last dimension by factor 2.

The *Discriminator*'s architecture starts from the image with dimension 64 × 64 × 1 and down-samples it to a tensor of size 4 × 4 × 512. This tensor is then flattened and fed through a linear layer outputting a scalar. 4 convolutional layers are followed by a *Leaky Linear Rectifier Units* layer.
Figure 6.9 gives an overview on the architecture of *Generator* and *Discriminator*.

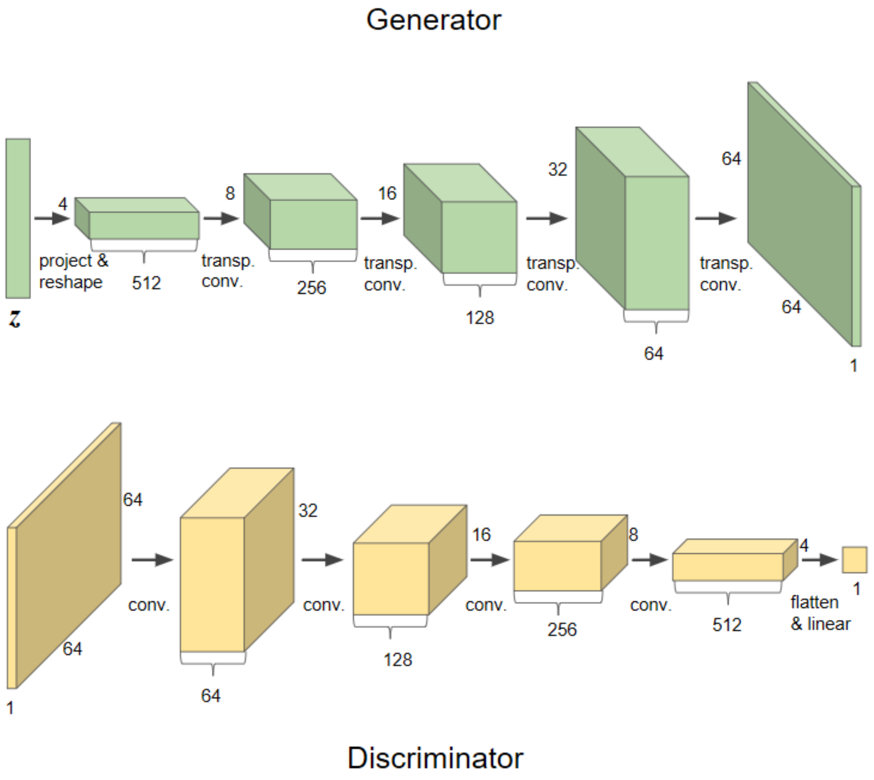

Figure 6.9: Overview of *Generator*'s and *Discriminator*'s architecture, figures adapted from [9]

The training procedure follows the suggestions in the Wasserstein-GAN paper [46], as described in Algorithm 4. The training images are scaled to lie in $[-1, 1]$ for training the GAN. At the beginning of the training and each epoch the data is randomly shuffled. Figure 6.10 shows images generated by the *Generator*.

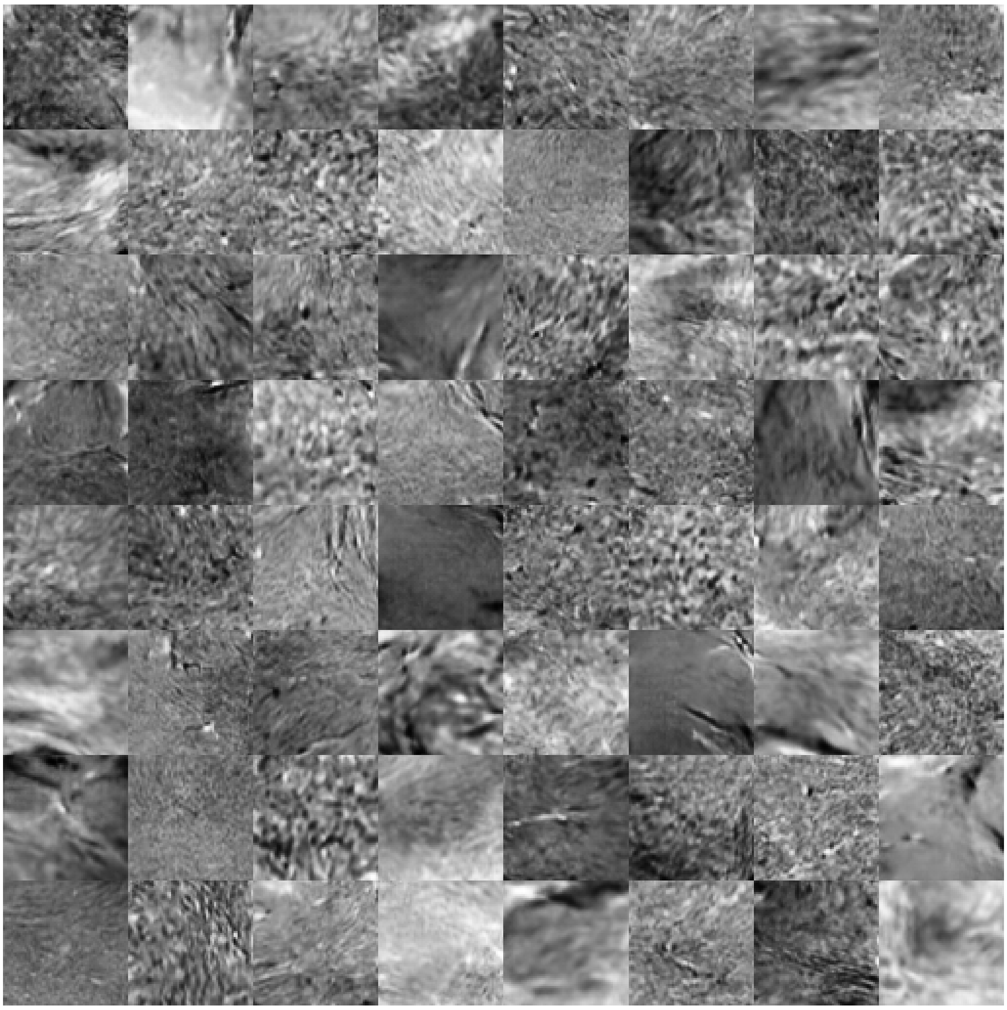

Figure 6.10: Examples of images created by the trained *Generator*

6.4.3 Down-sampling of New Images and Anomaly Score

After training is completed, the *Generator* has learned to reproduce patches showing healthy tissue. Images not seen during training are then mapped on the space of random noise input vectors and fed through the trained *Generator*, as in [28] and described in Section 4.3. The methodology of detection and prediction differs in term of the images which are down-sampled. For patients with lesions difference images of difference images $\boldsymbol{I}_{t_i-t_j}^{DiffOfDiff}$ are calculated by using the time point with confirmed lesion, $t_i = t_{lesion}$ as minuend in case of detection and one time point earlier, $t_i = t_{lesion-1}$ is used as minuend in case of prediction. More details on the composition of the evaluation sets are found in the next chapter.

The loss function used for mapping for detection as well as for prediction serves as anomaly score, however the loss function in this thesis is defined slightly different to the one given in Section 4.3, because the original scaling of the gray values is incorporated. They have been scaled to $[0, 250]$ before being saved as png. The loss function is given as

$$\mathcal{L}(\hat{\boldsymbol{z}}) = \lambda \mathcal{L}_R(\hat{\boldsymbol{z}}) + (1 - \lambda)\mathcal{L}_D(\hat{\boldsymbol{z}}), \tag{6.4}$$

with $\mathcal{L}_D$ as in 4.36 and $\mathcal{L}_R$ as

$$\mathcal{L}_R(\hat{\boldsymbol{z}}) = \sum |s_{orig}(\boldsymbol{x}) - s_{orig}(G(\hat{\boldsymbol{z}}))|, \tag{6.5}$$

where s_{orig} is the scaling to the original gray value range. Stochastic Gradient Descent with Nesterov Momentum is used to minimize the loss function in γ iterations. Afterwards the loss function serves as anomaly score. Figure 6.11 shows a score map for one slice of a patient with a lesion. The score map is created for a difference image $\boldsymbol{I}_{t_i - t_j}^{DiffOfDiff}$ with $t_i = t_{lesion}$ which means that a lesion, whose presence has been confirmed by an radiologist is detected. Nevertheless it serves as illustration of detection and prediction, since the score function is the same in both cases.

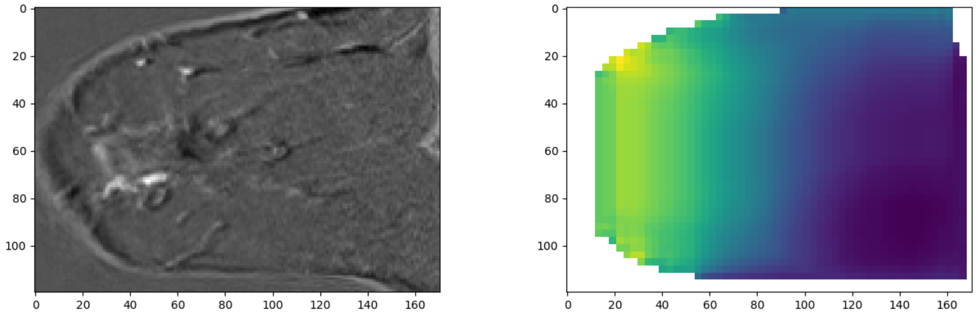

Figure 6.11: Score map for a patient with lesion (right) and reference image (left)

The maps have been calculated by moving a 64×64 size window over the slice to extract patches from top to bottom and left to right. After a patch has been extracted the window is moved one row downwards. For each pixel a mean score has been calculated by summing up the score values for the patches in which the pixel has been included and dividing by the number of visits of that pixel.

6.5 Discussion

In this chapter the methodological approach used in this thesis has been presented. After the preprocessing of the data the first step, which has been described is intra-subject image registration. The main difficulty caused the different resolutions of the images and that images acquired after 2014 show the breast with suppressed signal of fatty tissue, whereas the older images do not.

The next step has been the segmentation of the whole breast choosing the most similar breast model out of 9. Compared to Wengert *et al.* the reference images of each patient were not combined water-fat images as the model templates but DCE-MR images without fat suppression.

The last part has dealt with the anomaly detection. The approach here is as in [28], except the use of a Wasserstein GAN instead of the original proposed training procedure of a GAN in [70]. Besides, in this thesis the loss function used for mapping of new images includes the original scaling of the gray values in it definition.

CHAPTER 7

Experiments and Results

This part is devoted to the evaluation of the methods described in Chapter 6. Furthermore, experiments related to them are described. This chapter is mainly divided into five sections, whereas two discuss the registration methodology and one section the topics segmentation, anomaly detection and lesion prediction each. Although the main focus of this thesis lies on detection and prediction of breast lesions, registration and segmentation make up a necessary part of it without the detection and prediction were not possible in the way described in 6.4. Therefore, both deserve their own sections which discusses the topic.

The setup of the registration pipeline has been a step-wise process, which is described in Section 7.1. Different variants for the pipeline have been tested and are compared there. In Section 7.2 the registration pipeline is evaluated. Section 7.3 deals with whole-breast segmentation choosing the best breast model out of nine. The focus lies on the comparison of different templates representing the breast model and different registration strategies to match templates and an image belonging to a patient. The detection and prediction performance of breast lesions is investigated in Sections 7.3. and 7.4. In Section 7.3. different variants of training the *Generative Adversarial Network* and experiments concerning the choice of parameters for down-sampling are described, additionally. Each section closes with a discussion.

7.1 Initial Experiments to choose Registration Algorithm

The registration pipeline is, after preprocessing, the first method applied to the imaging data and has been built up stepwise. The setup process has included considerations concerning (a) the choice of the registration package, (b) putting an additional affine transformation prior to those calculated by Ezys and (c) how Ezys [55] calculates the non-rigid transformation. In the following, each of those considerations are described in more detail. Experiments are carried out for points (b) and (c). The aim has been to design a registration pipeline, which yields satisfactory optical results.

(a) Basic considerations when choosing the registration package The data set consists of 20 patients, however, at the beginning of the work it was assumed that 900 more patients would become part of the dataset. Therefore, finding a registration package, which is able to handle this amount of data in an acceptable amount of time has been the goal. This package has to be able to perform non-rigid registration to account for the change of breast shape over time [24]. Ezys is a reasonable choice for the purposes in this thesis, as it performs comparable to the SyN algorithm [42] but needing less time to match to images [55]. SyN achieved top rankings in the work of Klein *et al.* [58] in which the performance of different registration algorithms is compared.

(b) Advantages of an additional affine transformation The imaging data includes 3 different modalities and as an initial test, Ezys has been used to register two images acquired using different protocols. Troubles cause the registration of Modality 3 to one of the other 2 modalities and the registration the other way round, as the results are blurry images partly beyond recognition of the breast shape.
Ezys does perform an affine transformation before the non-rigid registration but it does not serve as a good initialization. The affine registration provided in the ANTs package [36] allows to easily customize parameters, such as the number of registration steps in each layer of a coarse-to-fine scheme. An affine transformation performed by ANTs initializes the non-rigid registration of a Modality 2 image to a Modality 3 image well enough, so as to obtain a breast shape which is not blurry.
Inspired by the fact that an additional affine transformation is necessary to achieve satisfactory results for the registration of Modality 2 and 3 images, this second transformation is carried out before every application of Ezys, with the hope to improve results compared to calculations, where an additional transformation is not performed.

(c) Variants to calculate non-rigid transformations Ezys offers the possibility to calculate the non-rigid transformation as the mean of forward and backward transformation. It is recommended to use this option, when one of the two images is interpolated [55], which is the case for inter-modal registration. Experiments and visual inspection investigate the usefulness of symmetric transformations.

7.1.1 Measures for Evaluation

Inspired by [64], for evaluation, the MI has been calculated after registration for every combination of modalities. For some matches, the dice has also been calculated, including registration of Modality 1 to Modality 2 and intra-modal registration of Modality 2 and 3. In case of intra-modal registration of Modality 3, the dice measures how well the glandular structures inside the breast match. Otherwise, the dice provides information how well the breast shapes in source and target image match after registration. It is calculated on binary images and lies between 0 and 1, whereas 0 means no match and 1 perfect match. A gray value image is transformed to a binary image using a threshold above which pixels are set to 1 and to 0 otherwise. This threshold is calculated using Otsu's method [51].

7.1.2 Data and Setup

MI and dice are calculated for 5 patients for every combination of Modalities. The pre-contrast images $\boldsymbol{I}_{t,m}^{pre}$ are used to calculate the transformations, however, it is refrained from the application of the transformations to the post-/pre-contrast difference images $\boldsymbol{I}_{t,m,pt}^{difference}$.
A direct registration of Modality 1 to Modality 3 images, even with an additional affine transformation, gives unsatisfactory results. However, matching Modality 1 and 2 images works. Therefore, Modality 1 images are registered to the last time point of Modality 2, $t = last_2$ at first, as pointed out in Figure 6.3. The reasons for registering Modality 2 images to $\boldsymbol{I}_{last_2,2}^{pre}$ at first instead to registering it to $\boldsymbol{I}_{last_3,3}^{pre}$ are that from a programming point of view, handling images of Modality 3 is done additionally to the steps which are carried out when those images are not available for a patient. The parts of the registration pipeline handling the registration within Modality 1 and 2 images are the same in both cases.

The metrics used for Ezys are *Normalized Mutual Information* and *Cross-Correlation. Mutual Information* is used for the affine ANTs - transformation. Besides, a 3-step course-to fine scheme has been chosen based on the recommendations in Avants *et al.* [36] with a maximum of 10,000 iteration steps in each of the three layers of the course-to fine scheme applied. ANTs takes the grid-size of target image as a reference for the number of pixels used in each step [2], therefore registering a low to a high resolution image is recommendable. This fact indicates registering older low resolution images to newer high resolution images and choosing the image frame of the chronologically last time point as target frame. As Ants, Ezys uses a coarse to fine scheme [36], [3]. The number of steps is set to 3 which is the default value [3]. In each step the algorithm subsamples twice [3]. The maximal grid size is set to 64 which is half of the default value [3].

7.1.3 Results

Additional affine transformation Table 7.1 bellow shows a comparison of *Mutual Information* values for matching two images, carried out with and without an affine ANTs-transformation. The mean values and the standard deviation are given.

Using *Cross-Correlation* to match Modality 3 images and images of other modalities does not work, since only for images of Modality 3 the signal of fatty tissue is suppressed. Registering Modality 2 and 3 without an additional affine transformation does not work either, as already discussed.
The versions of the pipeline with an additional transformation have higher values than the corresponding variants without. Compared to that, differences between the CC and NMI version with additional affine transformation are smaller.

Table 7.1: Mean values and standard deviation of Mutual information calculated for 5 patients

Modality	1 to 2	2 to 2	2 to 3	3 to 3	1 to 3
Add. trafo (nMI):	0.75±0.17	0.92 ± 0.15	0.34±0.08	0.49±0.27	0.11±0.04
No add. trafo(nMI):	0.64±0.13	0.82±0.12		0.46 ±0.27	0.095±0.03
Add. trafo (CC):	0.75±0.18	0.89±0.15		0.499±0.29	
No add. trafo(CC):	0.64±0.14	0.81±0.14		0.42±0.16	

For the dice, the numbers are, as for Mutual Information, better for versions with an affine transformation. Furthermore, results are better for the variants using Cross-Correlation as metric. The table bellow presents the results.

Table 7.2: Mean values and standard deviation of Dice Coefficient calculated for 5 patients

Modality	1 to 2	2 to 2	3 to 3
Additional trafo (NMI):	0.86±0.05	0.92 ± 0.01	0.77±0.05
No additional trafo(NMI):	0.8±0.06	0.90±0.01	0.78±0.03
Additional trafo (CC):	0.87±0.05	0.92±0.01	0.8±0.07
No additional trafo(CC):	0.85±0.04	0.91±0.02	0.78±0.05

Variants of nonrigid transformation. *Mutual Information* has been calculated for the symmetric variants of inter-modal registration. The values are lower than for the asymmetric variants. However, the advantage of using this option is also when comparing the images in Figure 7.1 below.

Table 7.3: Mean values and standard deviation of MI for symmetric transformation calculated for 5 patients

Modality	1 to 2	2 to 3	1 to 3
NMI:	0.72±0.16	0.32 ± 0.08	0.1±0.03

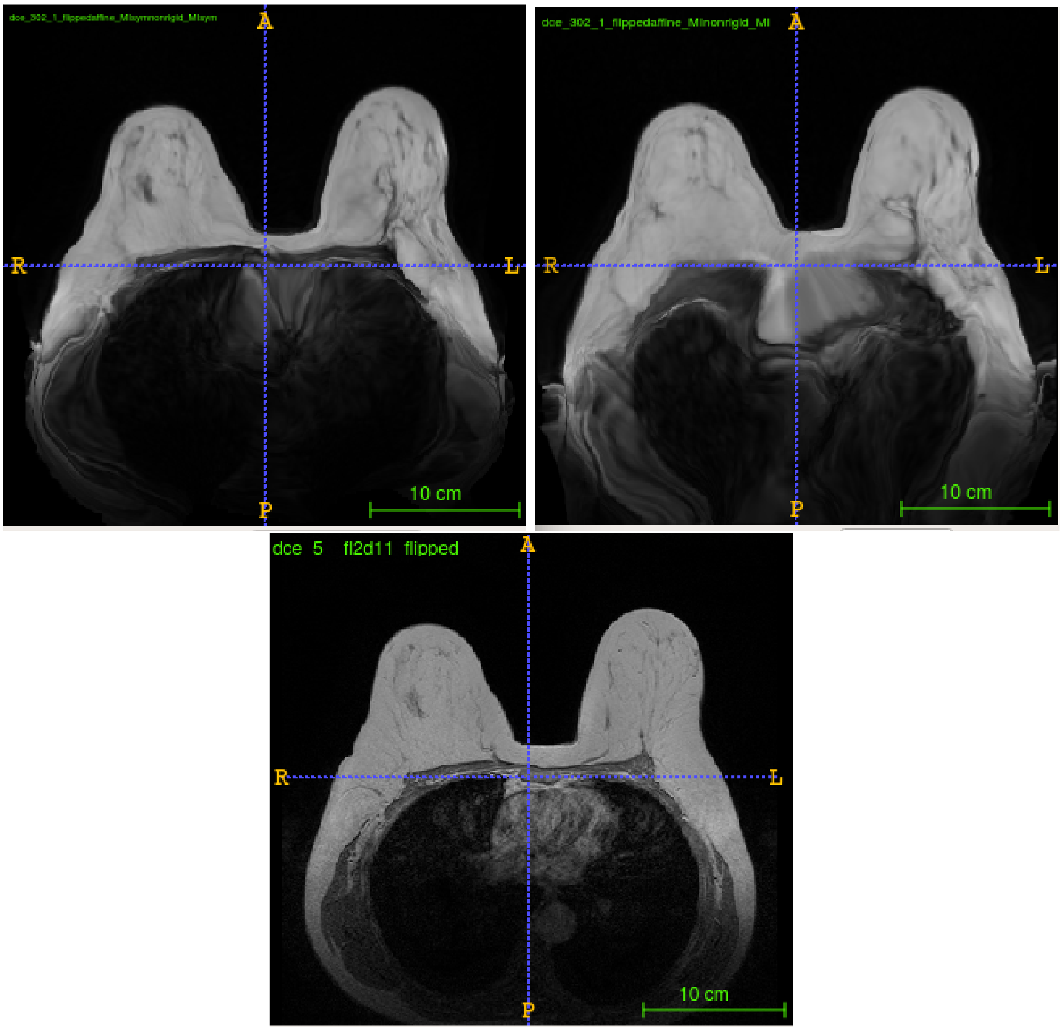
dce_302_1_flippedaffine_MIsymnonrigid_MIsym
A
R
L
P
10 cm
dce_302_1_flippedaffine_MInonrigid_MI
A
R
L
P
10 cm
dce 5 fl2d11 flipped
A
R
L
P
10 cm

Figure 7.1: Comparison of registration results: left: symmetric transformation, right: not averaged transformation, bottom: target image

In the right image parts of the pectoral muscle are used to match the heart in the target image. Averaging forward and backward transformation prevents this undesired behavior in the left image.

7.1.4 Discussion

In this section considerations and experiments concerning the setup of the registration pipeline have been described. The main difficulties have been caused by the difference in resolution of the 3 Modalities and by the fact that the signal of fatty tissue is only suppressed in Modality 3.

Results suggest, that additional affine transformation is useful as initialization for the non-rigid transformation applied to the data in this thesis. 2 similarity measures, namely CC and NMI have been tested at this step. As discussed in Chapter 3, *Crosscorrelation* is an intensity based metric, therefore voxels of the same or similar intensity are probably more likely to match in target and source image than for *Normalized Mutual Information*, resulting in higher values of the dice. For reasons of simplicity, however, *Normalized Mutual Information* is used for all possible combinations of modalities for source and target image.

Furthermore, averaging of forward and backward transformation is also used for intermodal registration as recommended in [55]. Figure 7.1 points out its usefulness, as the averaged transformation prevents the creation of a heart from tissue of the pectoral muscle. Building a heart makes sense from a computer's point of view, when using *Normalized Mutual Information* for matching, since in the target image, the heart is more visible than in the source image. Therefore, it is possible that the *Mutual Information* is higher for an image which is considered to be registered wrongly by an human observer than an image which looks well aligned but whose *Mutual Information* is lower.

7.2 Evaluation of registration

This section deals with the evaluation of the final registration pipeline, as described in 6.2. The aim is to evaluate how well two images match after registration.

7.2.1 Measures for evaluation

4 corresponding landmarks are set manually in pairs of source and target image. The transformed marked pixels get blurry through the non-rigid transformation applied, which results in a cluster of neighbored pixels being marked. For distance calculation the minimal euclidean distance between target landmarks and landmark cluster are calculated. The distances are also calculated by taking the pixel of a landmark cluster with the highest intensity value as a reference.

7.2.2 Data and Setup

The landmarks are set in 10 pairs of pre-contrast images $\boldsymbol{I}_{t,m}^{pre}$ which belong to 10 different patients and for each possible combination of target and source image 2 pairs are included in the overall set. The similarity measures used are MI for the ANTs-transformation and NMI for Ezys. Parameters, like the maximal grid size of Ezys or the number of iteration steps in the coarse-to fine scheme are the same as in the last section. 3 pairs require an initial alignment of the center of mass. As pre-contrast images $\boldsymbol{I}_{t,m}^{pre}$ and post-/pre-contrast difference images $\boldsymbol{I}_{t,m,pt}^{pre}$ are in the same frame for the same time point t and transformations are calculated based on images $\boldsymbol{I}_{t,m}^{pre}$, transformations are only applied to the pre-contrast images for evaluation.

7.2.3 Results

Table 7.4 gives a statistical overview of the minimal distances between landmarks in the target image and the corresponding landmarks in the transformed image. The values for the distances calculated by taking the pixel of a landmark cluster with the highest intensity value as a reference is given in Table 7.5.

Table 7.4: Statistical overview of minimal landmark distance between target and registered source image

mean	std	min	max	median
8.2626	29.9967	0	137.1209	0

Table 7.5: Statistical overview of landmark distance between target and registered source image based on maximal intensity

mean	std	min	max	median
10.8129	30.0482	0	139.3156	3.0602

The high values for mean an standard deviation are mainly caused by two landmarks, whose positions lie more than a 100 mm apart in source and target image. Those landmarks belong to the same patient, whose Modality 1 image has been registered to a Modality 3 image. The median shows that 50% of the landmarks have been positioned 3.0602 mm respectively 0 mm or less away from the corresponding landmarks in the target image.

The registration pipeline as described in Chapter 6 has been applied to 19 of 20 patients, delivering satisfactory results when observing them. The 20^{th} patient has only one time point available without breast implants, which makes this subject unusable for this work. Only 5 images which belong to different patients and modalities could not be registered without an initial alignment of the center of mass. For a 6^{th} case an initial alignment is not helpful. Since this time-point is the oldest available for the particular patient, it is excluded from subsequent work.

7.2.4 Discussion

This section has been devoted to the registration pipeline and its evaluation. Corresponding landmarks have been set in pairs of source and target image and the euclidean distance between them after registration has been calculated. The final pipeline gives overall satisfactory results when applied to 19 patients.

7.3 Evaluation of Segmentation

This part is devoted to experiments concerning the whole-breast segmentation. 2 different breast template types as well as 2 different variants to calculate the transformation required to match template and patient's breast are compared in terms of their usefulness for segmentation.

One of the two template types is the one used in Chapter 6 and is a combination of a water and a fat image. Using the decision function (6.2) the two images are combined to a water-fat template but it is also possible to obtain a fat-water template, if the positions of the water and fat image are interchanged in the decision function. This is the second template type. The differences of a water-fat and a fat-water image are pointed out in Figure 7.2.

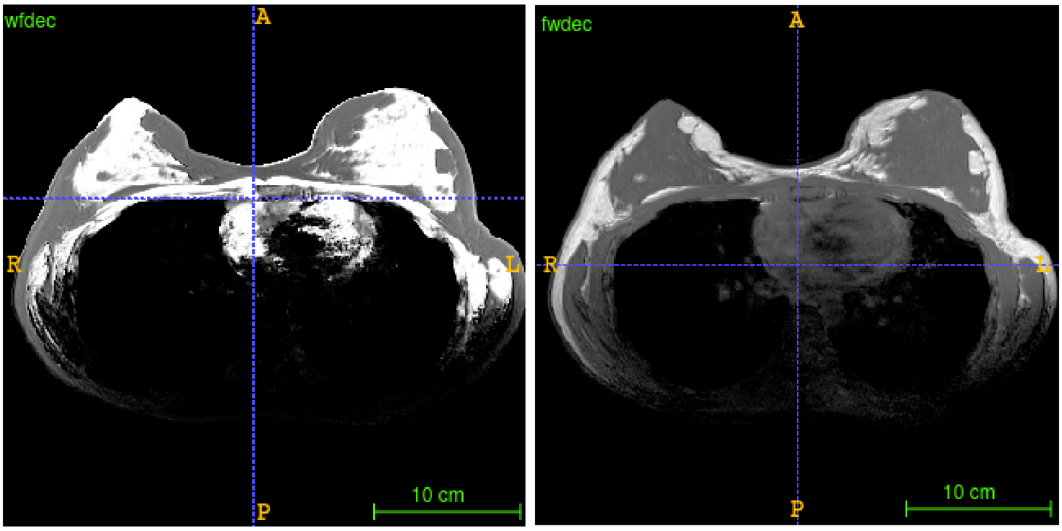

Figure 7.2: left: water-fat template; right: fat-water template

In water-fat templates fatty tissue has a lower intensity than tissue containing water. In fat-water templates it is the other way round.

The two variants to match patient's image and template are a nonrigid transformation with and without averaging forward and backward transformation. The first one conforms to the methodological approach in this thesis.

7.3.1 Measures for evaluation

The dice coefficient is calculated for a binary version of the patient's image $\boldsymbol{I}^{pre}_{last_2,m}$ and the registered breast template as in (6.2). The threshold to transform template and patients' images into their binary versions is set to 0.1.

7.3.2 Data and Setup

The evaluation is performed based on 5 patients. Mean and standard deviation are calculated for the dice values. Visual inspection of the segmentation performance is also part of the evaluation. Parameters for the affine ANTs-transformation and Ezys are the same as for the registration pipeline. The threshold used to choose the best breast model out of the nine during a segmentation process is 0.1.

7.3.3 Results

The table bellow summarizes the mean and standard deviation values of the dice coefficient calculated for 5 patients.

Table 7.6: Mean values and standard deviation of Dice coefficient

	water-fat template	fat-water template
sym trafo	0.9071±0.0197	0.9169±0.0110
no sym trafo	0.9164±0.011	0.9168±0.0169

Only the dice for the water-fat template combined with an averaged transformation differs from the other values in the second position after decimal point. Differences among the other 3 values are noise. Nevertheless, visually inspection of the 4 proposed segmentation procedures marks the use of water-fat-templates in combination with a single direction transformation as the best choice out of them. Figure 7.3 shows the segmentation obtained, when a fat-water template is used. The mask covers parts of the thoracic cavity which is not desirable.

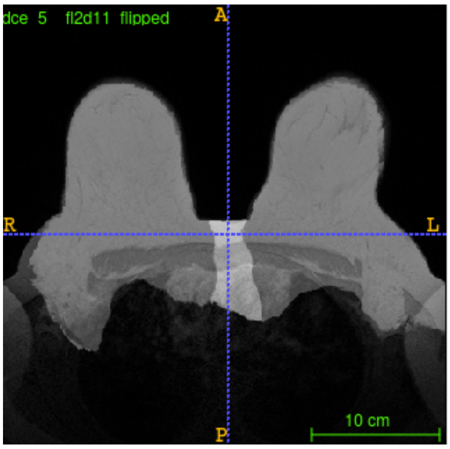

Figure 7.3: Segmentation performed using fat-water templates The mask covers also parts in the thoracic cavity.

The same problem for both template types appears if the non-rigid transformation is calculated as the average of forward and backward transformation. Figure 7.4 gives examples of falsely segmented breasts using a symmetric non-rigid transformation.

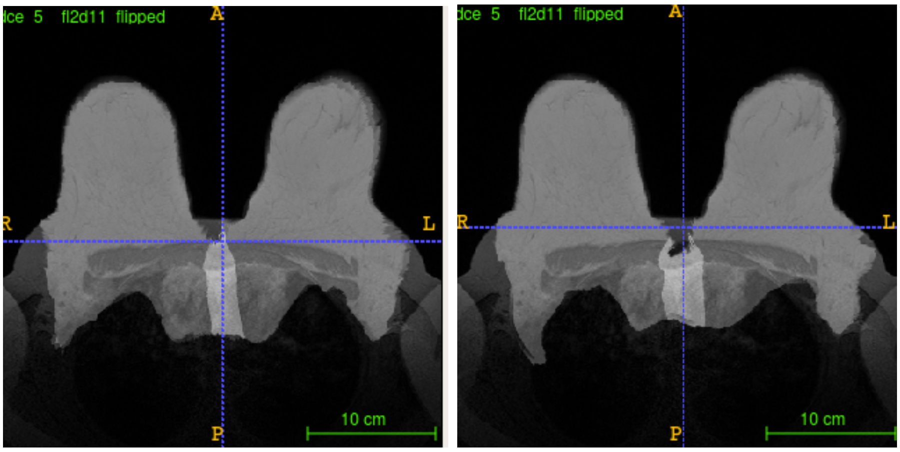

Figure 7.4: Segmentation performed using averaged transformation left: water-fat template; right: fat-water template

The problem that the mask also covers parts of the thoracic cavity has not been present when inspecting segmentation results for the 5 patients for evaluation, using the a water-fat template and transformation without averaging forward and backward transformation. The breast of the 19 patients, which are part of the data set of this thesis, have been segmented satisfactorily by the approach described in Chapter 6.

7.3.4 Discussion

Segmentation is necessary to be able to separate breast tissue from background and thorax pixels. 4 different variants to segment the breast have been investigated in this section, however all of them are based on choosing the most suitable breast model out of nine. This approach has been chosen because the breast models have already been available as well as code to determine the best model for a patient. In [45] a water-fat-image is calculated for each patient which is then used as a reference image. This is not possible for the data used in this thesis. Water-fat images are produced based on a T1 weighted image with fat suppression and a T2-weighted image. The latter are not available for each patient, but using $\boldsymbol{I}^{pre}_{last_2,2}$ as reference image for each patient also works.

One could assume that the fat-water templates are the better choice, because the intensities are visually similar tho those of $\boldsymbol{I}^{pre}_{last_2,2}$. However, the advantage of water-fat templates over the other template type is visually stronger separation from the thoracic cavity by the pectoral muscle and prevents the registration algorithm from pulling breast tissue inside the thoracic cavity.

Furthermore, averaging backward and forward transformation improved results of the registration pipeline and templates and patients' images do have different resolutions. However, in contrast to the registrations performed for spatial correspondence of time points, images with higher resolution are registered to images with lower resolution.

7.4 Evaluation of Anomaly Detection

In this section experiments concerning GAN-training and down-sampling of new images are described as well as the evaluation of anomaly detection methods in Section 6.4. One subsection is devoted to each of the three topics. Each closes with a discussion.

7.4.1 GAN-Training

The aim of experiments described in this section is to see if a GAN is able to learn a distribution of how breast tissue looks. Experiments concerning the training of the GAN include using the Wasserstein-GAN training procedure as well as the same as in the paper of Schlegl *et al.* [28]. Finding a training procedure and parameters which lead to a capture of the distribution of healthy tissue is desirable.

Measures of evaluation To evaluate the ability of the *Generator* to create patches showing breast tissue, samples of $\boldsymbol{z}$ have been draw randomly and fed though the trained *Generator*. The produced patches are then inspected. The sample size is 64 for each experiment, as for the face data set used in [9].

Data and Setup The experimental training set for the training procedure as in [28] includes a mixture of healthy and sick patients, as the information of the patients' health status was not available at that time. Furthermore, patches have not been extracted from difference images $\boldsymbol{I}_{t_i-t_j}^{diffOfdiff}$ of the first post-/pre-contrast difference images $\boldsymbol{I}_{t,m,2min}^{difference}$ but from $\boldsymbol{I}_{t,m,4min}^{difference}$, which have been acquired above 4 minutes after the injection of the contrast agent. $t_j = t_i - 1$ or $t_j = t_i - 2$ as described in 6.4.1. The size of the training set is 983.709.
Different values for the dimension of random input noise vectors $\boldsymbol{z}$ are tested. The parameter is set to $dim(\boldsymbol{z}) = 100$, $dim(\boldsymbol{z}) = 150$ and $dim(\boldsymbol{z}) = 200$. $dim(\boldsymbol{z}) = 100$ is used in Radford *et al.* [60], the other values are tested to see, if higher dimensions counteract mode collapse.
For the Wasserstein-GAN training procedure the random noise input vectors are an element of a 150-dimensional linear space. In all experiments 20 epochs are used, which is the same value as in [28].

Results Samples created by the *Generator* for $dim(\boldsymbol{z}) = 100$, $dim(\boldsymbol{z}) = 150$ and $dim(\boldsymbol{z}) = 200$ are shown in Figures 7.5-7.7. Regardless of the dimension of the $\boldsymbol{z}$-space, in all samples patches which look identical are found. This means that different points $\boldsymbol{z}$ are collapsed to the same image produced by the *Generator*. This issue is addressed by the Wasserstein-GAN. Therefore the training procedure has been changed to the one given in Algorithm 4. This eliminates the problem of mode collapse, which is visible in Figure 7.8, where all patches look different.

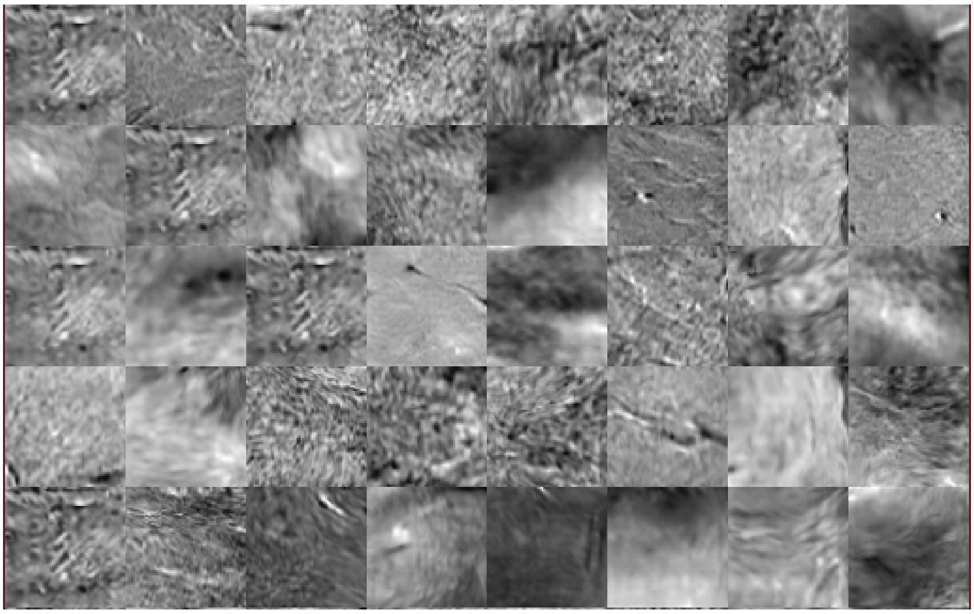

Figure 7.5: Generated samples with $dim(\boldsymbol{z}) = 100$ using the training procedure in [70]

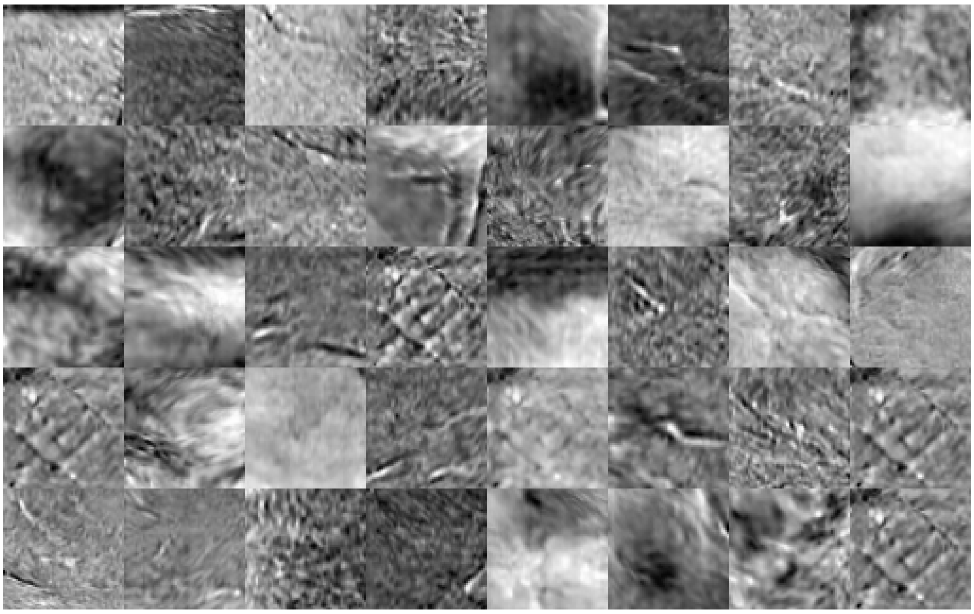

Figure 7.6: Generated samples with $dim(\boldsymbol{z}) = 150$ using the training procedure in [70]

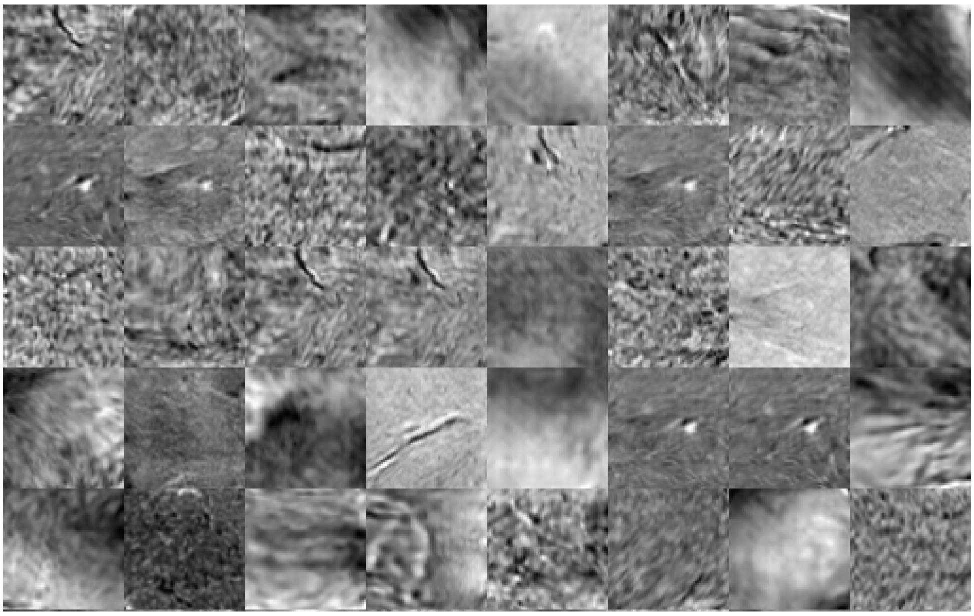

Figure 7.7: Generated samples with $dim(\boldsymbol{z}) = 200$ using the training procedure in [70]

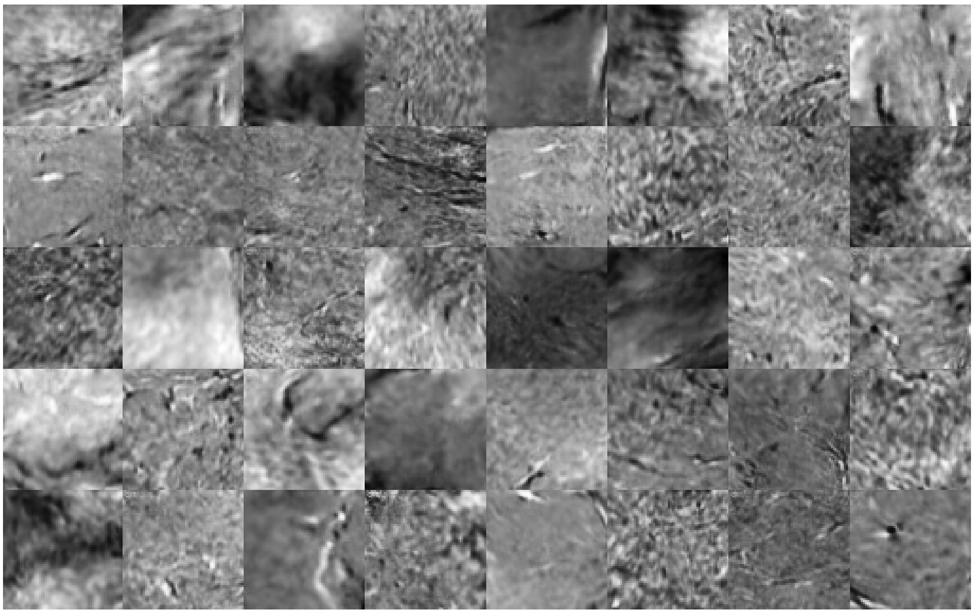

Figure 7.8: Generated samples with $dim(\boldsymbol{z}) = 150$ using the training procedure presented in Algorithm 4

Discussion In this part 2 different ways to train a GAN have been tested. Using the Wasserstein-GAN Training procedure with $dim(\boldsymbol{z}) = 150$ avoided the effect of mode collapse and is also used to train the GAN for evaluation of the detection and prediction performance.
At that point in the working process it is not relevant to distinguish between patients with and without lesions or which post-contrast images are a reasonable choice. The experiments keep their validity also for the training set, which is used for evaluation and includes only healthy patients.

7.4.2 Step-size for down-sampling of new images

The down-sampling of new images to the $\boldsymbol{z}$-space is done using a variant of gradient descent and therefore experiments are carried out to find a good step-size empirically. The experiments for the step-size are carried out for two different loss functions. The first one is the one proposed by Schlegel *et al.* [28] and described in 4.3. The second one is given in 6.4 and includes the original scaling of gray values for each patch. The usability of the two loss function for lesion detection is discussed in the next subsection.

Measures of Evaluation To find a good step size, different values are tested and the down-sampled images are compared to its original version. A step-size which is not too small is desirable for the sake of convergence speed.

Data and Setup The data set for finding a step-size consists of 64 patches extracted from healthy patients and 64 from patients with lesions. The gray values of the patches have been scaled to the interval $[0, 255]$ prior to saving them as png-files. They are taken out of the evaluation set, whose composition is explained in detail in the next subsection. The GAN has been trained on healthy patients using the Wasserstein-GAN training procedure with $dim(\boldsymbol{z}) = 150$. It is the same GAN as used in the evaluation process. More details on this are also found in the next subsection.

For the mapping which does not take the original scaling into account, step-sizes 0.01, 0.001 and 0.0001 and 0.00001 are tested, each with a number of up-dating steps of 500, 2,000 and 10,000. For those tests only the 64 patches of healthy patients are used. After a step-size has been chosen the patches with lesions are down-sampled with this step-size to see if it also works with these patches.
The experiments for the cost-function with original scaling of the gray values included have been carried out after those for the first cost-function. Knowledge obtained from the results of first experiments is incorporated, therefore step-sizes tested are 0.0001, 0.00001 and 0.000001 for 500 iterations. 0.000001 is also tested for 2,000 and 10,000 up-dating steps. The step-size is again chosen based on the 64 healthy patients and then used to down-sample the patches with lesions for checking the performance on those patches.

Results For the cost-function without scaling included 0.01 and 0.001 turn out to be too large. For 0.01 this is visible in Figure 7.9. 0.0001 and 0.00001 both work, however the larger step-size results in faster convergence, which is visible in Figure 7.10.

If the original scaling of the gray values is included in the cost-function, 0.0001 as well as 0.00001 turns out to be too large and is set to 0.000001 instead.

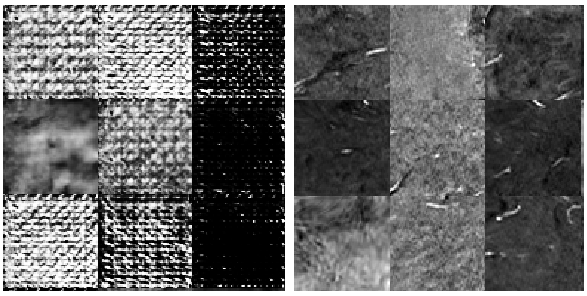

Figure 7.9: Generated samples with a step-size of 0.01 and 500 iterations (left) and original images (right), it is clearly visible that the step-size is too large.

Additionally to choose a step-size, qualitative results are given, which have been found when executing the experiments using the cost-function without scaling. Figure 7.10 shows the influence of increasing step-size when the number of up-date steps is kept the same, as well as the effect of increasing the number of iterations when the step-size is 0.0001. Figure 7.10 illustrates that the step-size has a larger impact on reaching convergence than increasing the number of up-dating steps.

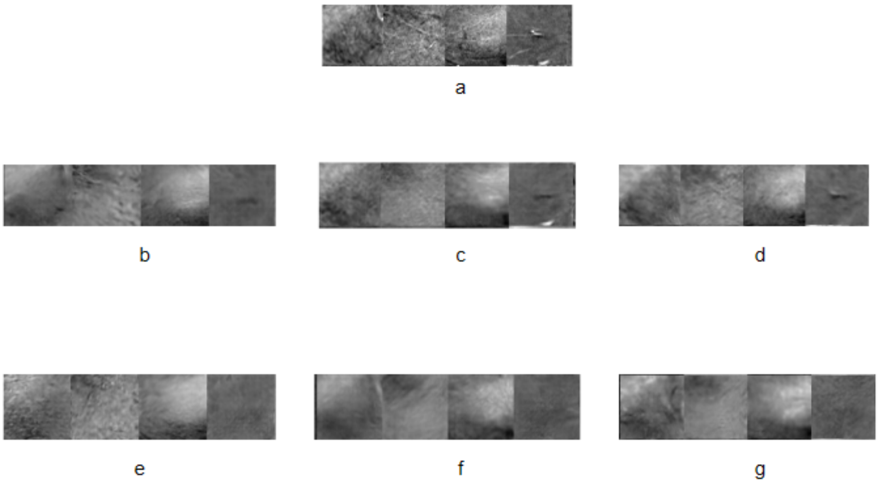

Figure 7.10: The figure illustrates the convergence of 2 step-sizes in comparison for different number of up-dating steps. The cost.function does not include the original scaling of gray values. **a**: original sample; **b-d**: step-size = 0.0001, number of iterations = 500, 2,000, 10,000; **e-g**: step-size = 0.00001, number of iterations =500, 2,000, 10,000;

When optimizing the values of $\boldsymbol{z}$ the corresponding generated image gets visually closer to the original image in terms of the dark and bright areas. This is not only the case for patches showing healthy tissue as in Figure 7.10 but also for patches which include lesions, as in Figure 7.11, **a-c**.

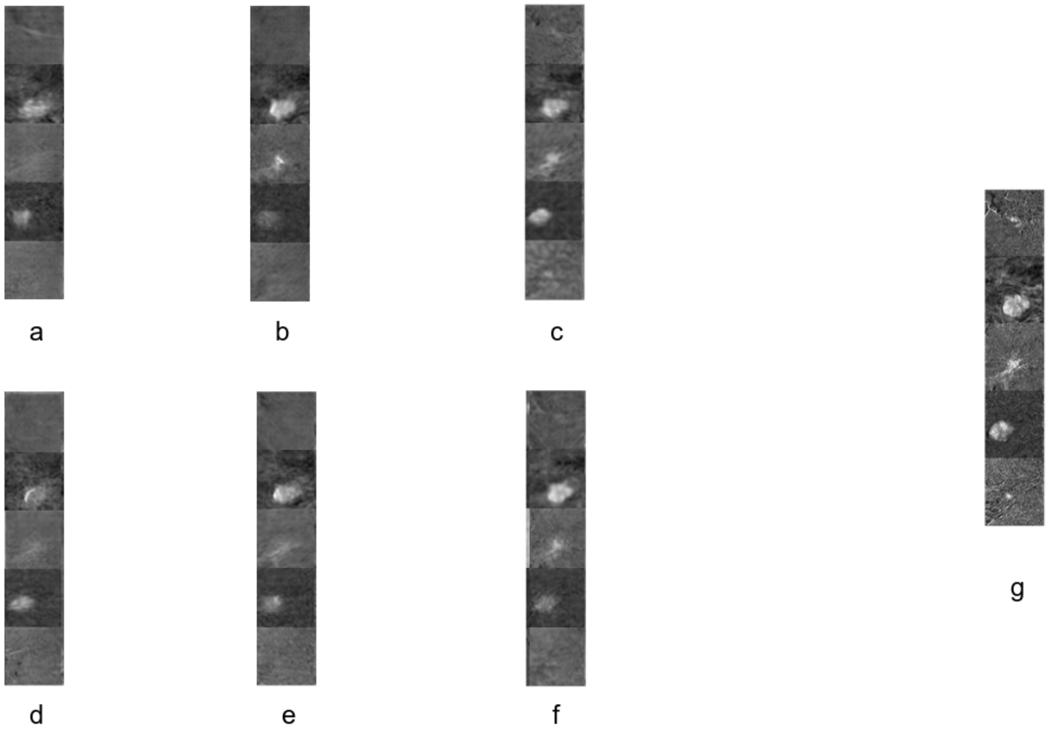

Figure 7.11: Illustration of results of patch creation for patches containing lesions. The *Generator* is able to rebuilt patches showing lesions. **a-c**: original gray value scaling not included, step-size = 0.0001, number of iterations = 500, 2,000, 10,000; **d-f**: original gray value scaling included, step-size = 0.000001, number of iterations = 500, 2,000, 10,000; **g**: original sample

If the original scaling of the gray values is included in the cost-function, the qualitative behavior is the same concerning creation of patches containing lesions as without including the original scaling, which is also visible in Figure 7.11.

Discussion In this subsection experiments have been executed to find a step-size for the down-sampling of new images. Two cost-functions have been investigated, one with and one without the original scaling of gray values included. The step-size choices for both variants are different, however the qualitative behavior is similar. In the next subsection it will be visible that including the original scaling of gray values has an impact on the anomaly detection performance.

7.4.3 Evaluation

In this final subsection the detection ability of the approach described in 6.4 is evaluated concening its ability to distinguish between patches containing a lesion and those which does not. Differences between malign and benign lesions are also discussed. Additionally, the performance using the cost-function without the original scaling of gray values included is also investigated.

Measures for evaluation Receiver Operator Characteristics (ROC) curves are calculated for evaluation. For a ROC curve the anomaly scores for each patch are calculated first. Then the classification threshold above which a patch is classified as containing a lesion is set equal to each of the obtained scores one after the other and corresponding *True Positive Rates* and *False Positive Rates* are calculated. The *True Positive Rates* on the second axis are then plotted over the *False Positive Rates* on the first axis.
Positive refers to patches which contain lesions whereas *Negative* refers to patches showing only healthy tissue. A *True Positive* (TP) is a patch containing a lesion which is also classified as such. A *False Positive* (FP) is a patch without lesion, which is incorrectly classified as *Positive.* The *True Positive Rate* is the fraction of Positives, which are correctly classified and the *False Positive Rate* is the fraction of Negatives which are wrongly classified as Positive. Equivalent terms exist for *Negatives.*

As measures the area under the curve is calculated as well as Accuracy, Precision, Sensitivity and Specificity for the best split according to the Youden's index. The area under the curve is the area which is bounded by the first axis and the ROC curve. The Youden's index is used by Schlegl *et al.* [28] to determine the best splitting value for a particular ROC curve. For each point of the ROC curve the Youden's index Y is calculated by

$$Y = Sensitiviy + Specificity - 1. \tag{7.1}$$

The point with maximal index is then chosen as best splitting point. Equations for Sensitivity, Specificity, Accuracy and Precision are summarized in Table 7.7.

Table 7.7: Equations for Accuracy, Precision Sensitivity and Specificity, N is the data size

Accuracy	Precision	Recall/Sensitivity	Specificity
$\frac{TP+TN}{N}$	$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FN}$	$\frac{TN}{TN+FP}$

Data and Setup The data set used to evaluate the methodology described in Chapter 6 originally consists of 10 patients who have lesions either malign or benign and 10 patients without lesions who are considered to be healthy. However, 2 patients have to be excluded completely as well as single time-points belonging to different patients. One reason for the exclusion of single time-points is the complete loss of the breast or the insertion of breast implants. This applies to 3 patients and has caused one to be excluded completely. The second reason has been wrong registration, which happened to one time-point. The complete exclusion of the second patient has been decided during the evaluation process of the anomaly detection abilities of the approach. To be useful for evaluation a patient needs time-points without lesions and this property the particular patient does not have.

The GAN is trained on patches are extracted from 5 healthy patients as described in Chapter 6. 20 epochs are used for training. The extraction procedure is carried out on the difference images $\boldsymbol{I}_{t_i-t_j}^{diffOfdiff}$ of first post-/pre- contrast difference images $\boldsymbol{I}_{t,m,2min}^{difference}$ and creates a training set of 1,712,238 patches. These patches look similar when extracted from the same slice with translation of the structures visible being the only difference. As discussed in Chapter 2 the first post-contrast difference images $\boldsymbol{I}_{t,m,2min}^{difference}$ are useful to detect malign lesions. This type of lesions appear in 5 patients whereas benign lesions appears in 3.

For the evaluation of the anomaly detection performance based on the trained GAN, a test set consisting of extracted patches from the 8 not excluded patients with lesions and the five healthy patients not used for training is created.
For healthy patients, the first post-contrast difference images $\boldsymbol{I}_{t,m,2min}^{difference}$ of a time point are subtracted from the first post-contrast difference images $\boldsymbol{I}_{t_n,m,2min}^{difference}$ belonging to the youngest time point available for a patient, with t_n time points available. The time points which are subtracted are either acquired by the same modality as images belonging to the chronologically last time point t_n or by the modality which was used before (compare Table 6.1 for information on the temporal usage of modalities). Therefore, images of Modality 1 and 3 are not subtracted from each other. The reason to exclude those images lies in the fact, that the GAN has not seen difference images of difference images between Modality 1 and 3.
For patients with lesions the difference images are calculated the same way but the youngest time point is chosen as the time point $t_{lesions}$ where the lesion has been detected by a radiologist the first time.
After the calculation of difference images, patches are extracted slice-wise allowing 50 pixels outside the breast mask. Before saving the patches, the gray values have been scaled to the range 0-255. The extraction procedure gives 38637 patches which show healthy tissue and 1056 patches containing lesions. 64 patches of both classes are taken out and used for the experiments in the previous subsection. Therefore, the test set consists of 992 patches containing lesions and 38573 patches of healthy patients.

ROC curves are calculated for the score version, which does not include the original scaling of the gray values for 500, 2,000 and 10,000 up-dating steps. The step-size is set to 0.001. 128 patches are drawn randomly out of each of the pools of healthy patches and those showing lesions.

For the anomaly score with the original scaling included the same has been done except that the step-size is set to 0.000001. Additionally, ROC curves are calculated 5 times more on five different randomly sampled subsets for 2,000 up-dating steps, each containing 128 examples of lesions and 128 healthy patches. Based on these 5 sets a mean value for the evaluation measures are calculated.

Results Regardless, which number of iterations is chosen for the cost-function without including the original scaling of gray values, the ROC curve is slightly S-shaped. This is visible in Figure 7.12. If the achieved scores of each patch involved in the calculation of the ROC curves are ordered from low to high values, patches with scores belonging to the last quarter are classified as Positive although they are not. This is visible as the ROC curve lies bellow the line representing random guessing in the first quarter of the figure. This means that patches which show healthy tissue are classified as containing a lesion and that there are patches which do not contain lesions achieve a higher score than patches which do contain lesions. This violates the assumption that patches showing only healthy tissue are scored lower than those with lesions.

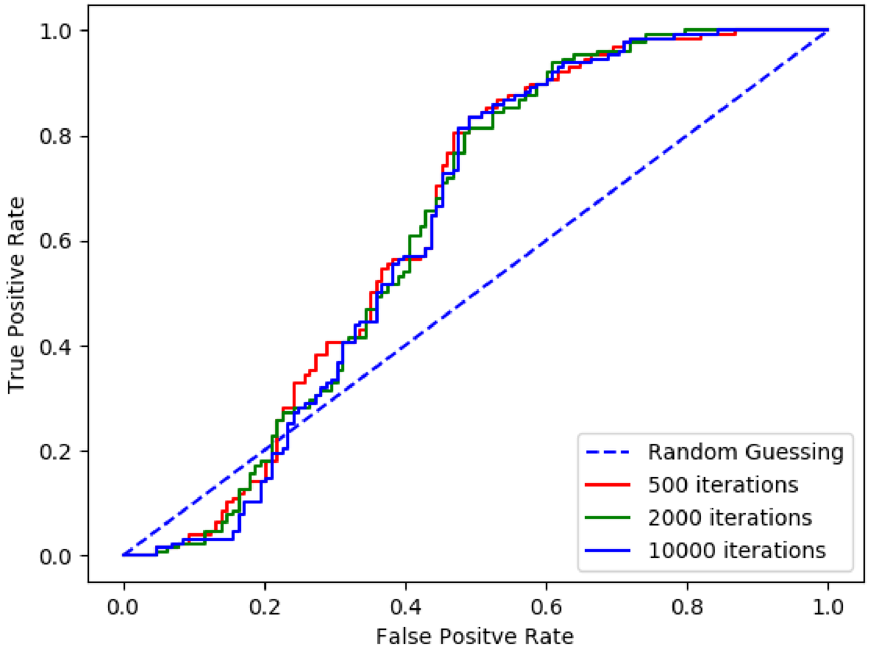

Figure 7.12: ROC curve for anomaly score without original intensity scaling included

The anomaly score with the original scaling included gives better results. One example for a ROC curve is given in Figure 7.13.

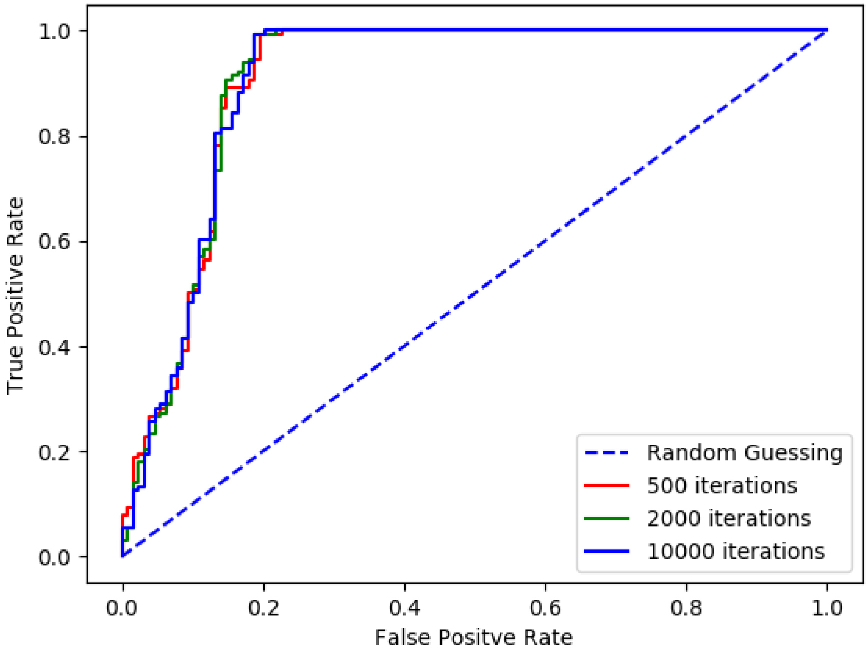

Figure 7.13: ROC curve for anomaly score with original intensity scaling included

The area under the curve is 0.9 for each number of iterations and the *True Positive Rate* is 0.99 at the best splitting value according to Youden's index, whereas the false positive rate is 0.2 for 500 iterations and 0.19 else. Table 7.8 summarizes some performance measures calculated for each of additional 5 subsets.

Table 7.8: Detection: Performance measures for 5 different subsets calculated for best split according to Youden's index

	Accuracy	Precision	Recall/Sensitivity	Specificity	AUC
Subset 1	0.945	0.901	1	0.891	0.95
Subset 2	0.922	0.865	1	0.844	0.91
Subset 3	0.902	0.846	0.984	0.820	0.88
Subset 4	0.914	0.858	0.992	0.836	0.91
Subset 5	0.906	0.842	1	0.813	0.87
Mean	0.918	0.862	0.995	0.841	0.90

On average more than 99% of patches containing lesions are correctly identified as such and less than 20% of patches showing healthy tissue are falsely classified as anomalous. This is visible from Table 7.8, since $False\ Positive\ Rate = 1 - Specificity$.

Since the first post-contrast difference images are especially helpful to detect malign lesions [33], it is also investigated for what kind of patches the detection of present lesions fails. It turns out that all those patches belong to the same patient. This subject has lesions classified as malign. This patient does not have one single lesion but multiple lesions which are distributed in her luctiferous ducts. Patches which were classified wrongly, contain 3 pixels or less which are segmented as lesion. Patches belonging to the same patient and containing more tissue marked as lesion have not appeared as a *False Negative*.

Finally, the results presented in the previous subsection are compared to those achieved in the paper of Schlegl *et al.* [28] and those in the paper of Hadad *et al.* [26].
The approach described in the former paper has inspired this work and is applied to images of the retina. An area under the curve of 0.89 has been achieved on a test set consisting of 8,192 image patches extracted from 20 patients. 10 of those patients are considered to be healthy, the other 10 not. Values for Precision, Sensitivity and Specificity are 0.883, 0728 and 0.893. The values of Precision and Specificity are higher than the corresponding mean values in Table 7.8, however the Sensitivity is lower.
Hadad *et al.* train a CNN on MR imaging data to classify mass and non-mass regions in the images. Data of 123 patients is included in the data set. For each patch containing a lesion, a patch of similar size is extracted showing normal tissue. Afterwards the extracted patches are scaled to the input size needed by the CNN. Through augmentation the data set is enlarged to 19,316 images. The performance has been evaluated using Monte Carlo Cross-Validation with 50 repetitions and a splitting ratio of 90% for training and 10% for evaluation. An area under the curve of 0.98 and an accuracy of 0.94 have been achieved. Both values are higher than the corresponding values in Table 7.8.

Discussion The anomaly detection performance of the approach described in 6.4 has been evaluated and the detection performance using the cost-function in [28] has been investigated. This investigation delivered worse results than the approach of this thesis. The results of the methodology described in this thesis also have been compared to those in [28] and [26].
The evaluation set used in this thesis includes the smallest amount of patients compared to [28] and [26]. Furthermore, the number of patients included in the GAN-training set is, with 5 patients included, also smaller than the 270 patients who contributed to the GAN - training set in [28]. Although the 20 patients provided for this thesis are randomly chosen from the amount of patient data the Vienna General Hospital possesses, evaluating the approach on more patients makes the results more reliable. The pool of healthy patches and patches showing lesions contains groups of images which are very similar and differ only by translation of the structure visible in them. To evaluate the performance on patches showing different structures and lesions, a subset has been selected randomly

several times. Based on these subsets mean values for different performance measures have been calculated. Cross validation, as in [26] has not been carried out, as the training of the GAN takes 3 days. The values obtained for accuracy, AUC, precision, recall and specificity are comparable to those obtained in Schlegl *et al.* [28], however the examined medical structure is different. The approach in [26] achieved better results in terms of AUC and accuracy than the one described in this thesis. But a conceptual advantage of the methods in this work is (theoretically) the detection of everything that diverges from the distribution of healthy tissue, whereas Hadad *et al.* learn to distinguish between mass and non-mass lesion and not all breast lesions are mass regions [23].

7.5 Evaluation of Anomaly Prediction

This final section is devoted to the evaluation of the prediction performance. Since the prediction methodology is similar to detection, the evaluation of both is also similar.

7.5.1 Measures for evaluation

ROC curves, area under the curve and other performance measures based on the Youden's index are calculated analogous to the evaluation of the detection performance.

7.5.2 Data and Setup

The GAN has been trained on the same training set and with the same parameters as for detection. The test set to evaluate how well lesions are predicted is built up similar to the test set for detection, as described in the last section. The only difference is that for patients with lesions, the time point $t_{lesion-1}$ before the lesion has been detected by a radiologist has been chosen as minuend. This has given 546 patches from 7 patients. For the eighth patient only two time points are available overall, therefore it has been excluded from the evaluation of the prediction performance.

As for detection, for the cost-function with original scaling included as in 6.4.3, ROC curves are calculated on 5 subsets with 128 patches drawn randomly from each of the two pools. The step-size for down-sampling is set to 0.000001.

7.5.3 Results

Results are comparable to the detection performance and are visible in Figure 7.14 and Table 7.9.

In Figure 8.14 the AUC is 0.90 for each number of iterations. The *True Positive Rates* are 0.91, 0.94 and 0.94 and the *False Positive Rates* are 0.17, 0.20 and 0.18. According to Table 7.8 the prediction performance is on average lightly worse than the detection performance. *False Negatives* are again patches only extracted from the patient already

Table 7.9: Prediction: Performance measures for 5 different subsets calculated for best split according to Youden's index

	Accuracy	Precision	Recall/Sensitivity	Specificity	AUC
Subset 1	0.867	0.836	0.914	0.820	0.89
Subset 2	0.875	0.804	0.992	0.758	0.87
Subset 3	0.855	0.794	0.960	0.75	0.88
Subset 4	0.816	0.779	0.883	0.75	0.83
Subset 5	0.867	0.856	0.883	0.852	0.90
Mean	0.856	0.814	0.927	0.786	0.87

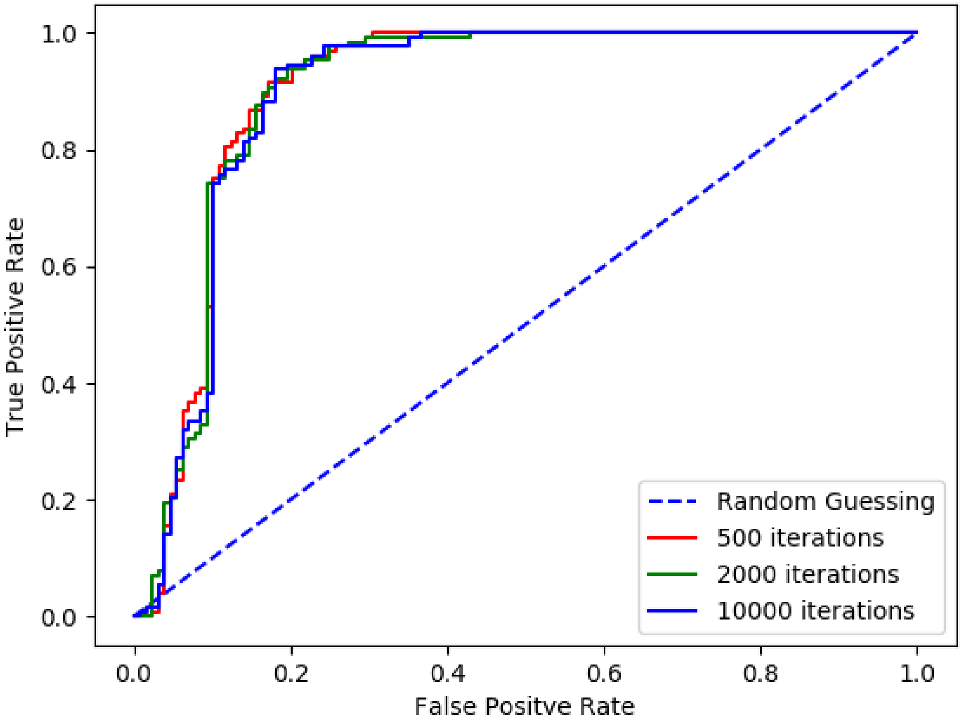

Figure 7.14: ROC curve for prediction performance

mentioned. Again, in all patches which are wrongly classified as healthy, not more than three pixels are marked as anomalous.

7.5.4 Discussion

In this section the prediction performance of the methodology in 6.4 has been discussed. With the approach described, it is possible to identify lesions before they where confirmed by a radiologist. As been told by a radiologist, if a lesion is confirmed in an actual acquisition series retro-perspective investigation of older acquisition time points sometimes shows that the lesion has been present before. The approach discussed in this thesis may therefore be used as guide post to point at regions in breast tissue, which may contain already a lesion.

CHAPTER 8

Conclusion

This chapter constitutes a recapitulation of the key points of this thesis. Furthermore, thoughts about possible future work are also given.

8.1 Summary

The aim of this thesis is the development and application of a GAN-based score to detect lesions in DCE-MR images of high risk patients for breast cancer, as well as the evaluation of its performance as anomaly predictor. The latter corresponds to the detection of a lesion by the GAN at an earlier stage than a radiologist.
The main steps to reach the goals of this work are intra-patient registration, whole breast segmentation and finally the anomaly detection. Registration has been necessary to reach spatial correspondence between all images belonging to the same patient, since several time points of acquisition have been available for each subject. Matching of images has been performed by the diffeomorphic demons algorithm Ezys [55] and two prior affine transformations. One of those is calculated automatically when running Ezys but did not form an initial point good enough for the nonrigid transformation. The other one has been added to obtain a good initialization and is carried out by the ANTs package [**?**].
The next corner stone, which is the segmentation, has been calculated by choosing one breast model out of 9, whose size and shape fits the patient's data best after nonrigid registration. Each breast model consists of a template and the corresponding segmentation. The template is a combined water-fat image in which fatty tissue appears darker than tissue which contains water.
The final part is the anomaly detection itself, carried out using a GAN-based anomaly score. The GAN is trained to learn the distribution of patches showing healthy tissue. Then new images are down-sampled to the input space of the *Generator*, so as to generate an artificial version coming as close as possible to the original image. The anomaly score represents how far the generated image is away from the original image, whereas healthy

tissue has a lower score than patches containing a lesion.
The main contribution of this work has been the use of temporally linked imaging data for anomaly detection and a GAN-based anomaly score for detection of malign and benign breast lesions. The method has shown to be able to detect lesions earlier than a radiologist.
To conclude, the proposed approach for breast lesion detection shows promising results on the available data set. However, an evaluation on more than 13 patients (5 healthy, 8 with lesions), would give more confidence in the performance of the method.

8.2 Future Work

The presented lesion detection approach offers possibilities for future work and improvement. One is the evaluation on a larger data set. Other possibilities are listed bellow.

Combination of DCE-MRI with other modalities The sensitivity of the described method is high with a mean value of 99.5 %, however the specificity, which is 84%, might increase, if additional MR imaging techniques, such as Diffusion Weighted Imaging (DWI), serve as information source. Multi-modal breast imaging is already in use, whereas DCE-MRI serves as backbone of an imaging protocol and adding further modalities increases specificity [22]. One possible implementation of the combined usage of different modalities could be to calculate a GAN-based score for each modality and obtain a combined score as a weighted sum of the single scores.

Classification benign vs malign Investigating if the anomaly score can be adapted in some way, so as it is possible to classify lesions into benign an malign is another interesting possibility for future work. In this thesis the first post-contrast images have been used for evaluation and training of the GAN. One first step in the direction of classification could be to exchange the first post-contrast images with post-contrast images acquired 4 or 6 minutes after the injection of the contrast agent and evaluate the anomaly detection performance. The latter is more suitable to detect benign lesions, whereas the first post-contrat images are better to identify malign lesions [33]. Probably it works to calculate an anomaly score for the first post-contrast images and for those acquired after 4 or 6 minutes at first. Then, if the score for the first post-contrast images is higher the lesion is classified as malign and if the other score is higher, it is classified as benign.

Bibliography

[1] https://github.com/rordenlab/dcm2niix
https://www.nitrc.org/plugins/mwiki/index.php/dcm2nii:MainPage.

[2] Anatomy of an antsregistration call. https://github.com/ANTsX/ANTs/wiki/Anatomy-of-an-antsRegistration-call, last accessed on 20.07.2018.

[3] Ezys image registration. https://www.yumpu.com/en/document/view/3994074/ezys-image-registration-bss-university-of-cambridge/7, last accessed on 20.07.2018.

[4] B. T. Polyak . Some methods of speeding up the convergence of iteration methods. *Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

[5] V. Giannini; A. Vignati; L. Morra; D. Persano; D. Brizzi; L. Carbonaro; A. Bert; F. Sardanelli; D. Regge. A Fully Automatic Algorithm for Segmentation of the Breasts in DCE-MR Images. *32nd Annual International Conference of the IEEE EMBS*, 2010. DOI: 10.1109/IEMBS.2010.5627191.

[6] D. Selvathi; A. AarthyPoornila. Performance analysis of various classifiers on deep learning network for breast cancer detection. *International Conference on Signal Processing and Communication (ICSPC'17)*, 2017. DOI: 10.1109/CSPC.2017.8305869.

[7] J. Levman; T. Leung; P. Causer; D. Plewes; A. L. Martel. Classification of Dynamic Contrast-Enhanced Magnetic Resonance Breast Lesions by Support Vector Machines. *IEEE Transactions on Medical Imaging*, 27(5):688 – 696, 2008. DOI: 10.1109/TMI.2008.916959.

[8] A. L. Maas; A. Y. Hannun; A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. *ICML Workshop on Deep Learning for Audio, Speech, and Language Processing*, 2013. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.693.1422&rep=rep1&type=pdf, last accessed on 29.08.2018.

[9] Brandon Amos. Image Completion with Deep Learning in TensorFlow. `http://bamos.github.io/2016/08/09/deep-completion`. Accessed: [13.06.2017].

[10] A. Gubern-Mérida; R. Martí; J. Melendez; J. L. Hauth; R. M. Mann; N. Karssemeijer; B. Platel. Automated localization of breast cancer in DCE-MRI. *Medical Image analysis*, 20:265–274, 2015.

[11] J. Yao; J. Chen; C. Chow. Breast Tumor Analysis in Dynamic Contrast Enhanced MRI Using Texture Features and Wavelet Transform. *IEEE Journal of Selected Topics in Signal Processing*, 3(1):94–100, 2009. DOI: 10.1109/JSTSP.2008.2011110.

[12] C. E. Shannon;. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423/623–656, 1948.

[13] D. L. G. Hill; D. J. Hawkes; N. A. Harrison; C. F. Ruff. A strategy for automated multimodality image registration incorporating anatomical knowledge and image characteristics. *Information Processing in Medical Imaging*, 687:182–196, 1993.

[14] St. Marrone; G. Piantadosi; R. Fuscoy; A. Petrilloy; M. Sansone; C. Sansone. Breast segmentation using Fuzzy C-Means and anatomical priors in DCE-MRI. *23rd International Conference on Pattern Recognition*, 2016. DOI: 10.1109/ICPR.2016.7899845.

[15] S. Ioffe; Ch. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 2015. arXiv:1502.03167v3 [cs.LG] https://arxiv.org/abs/1502.03167, last accessed 29.08.2018.

[16] D. Rueckert ; L. I. Sonoda ; C. Hayes ; D. L. G. Hill; M O Leach; D J Hawkes . Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging*, 18(8):712–721, 1999.

[17] M. Kim, G. Wu; D. Shen. Groupwise registration of breast DCE-MR images for accurate tumor measurement. *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 598–602, 2011. DOI: 10.1109/ISBI.2011.5872478.

[18] W. Bidgood Jr.; S. Horii; F. Prior; D. Van Syckle. Understanding and using DICOM, the data interchange standard for biomedical imaging. *Journal of the American Medical Informatics Association*, 4(3):199–212, 1997.

[19] C. Studholme; D.L.G. Hill; D.J. Hawkes. Multiresolution voxel similarity measures for MR-PET registration. *Information Processing in Medical Imaging*, pages 287–298, 1995.

[20] C. Studholme; D.L.G. Hill; D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.*, 32(1):71–86, 1999.

[21] O. Dössel. *Bildgebende Verfahren in der Medizin*, chapter Magnetresonanz-Tomographie, pages 285–390. Springer-Verlag Berlin Heidelberg, 2016. DOI: 10.1007/978-3-642-54407-1_11.

[22] K. Pinker; T. H. Helbich; E. A. Morrris. The potenital of multiparametric MRI of the breast. *The British Journal of Radiology*, 90(1069), 2016. https://doi.org/10.1259/bjr.20160715.

[23] N. Sakamoto; M. Tozaki; K. Higa; Y. Tsunoda; T. Ogawa; S. Abe; S. Ozaki; M. Sakamoto; T. Tsuruhara; N. Kawano; T. Suzuki; N. Yamashiro; E. Fukuma. Categorization of non-mass-like breast lesions detected by mri. *Breast Cancer*, pages 241–246 Volume 15(3), 2014. Online ISSN: 1880-4233.

[24] F. Pera E. T. Peuker, T. J. Filler. *Waldeyer Anatomie des Menschen*, chapter Brustkorb, Thorax und Brustraum, Cavitas thoracis mit Zwerchfell, Diaphragma, pages 781–902. Fanghänel J., Pera F., Anderhuber F., Nitsch R. (Ed.), Walter de Guyter, Berlin, 17^{th} edition edition, 2003. online e-book, accessed 25.04.2017 https://www.degruyter.com/viewbooktoc/product/42130.

[25] V. Dumoulin; F. Visin. A guide to convolution arithmetic for deep learning. 2018. arXiv:1603.07285v2 [stat.ML] https://arxiv.org/abs/1603.07285 last accessed on 29.08. 2018.

[26] O. Hadad; R. Bakalo; R. Ben-Ari; Sh. Hashoul; G. Amit. Classification of breast lesions using cross-modal deep learning . *2017 IEEE 14th International Symposium on Biomedical Imaging*, 2017. DOI: 10.1109/ISBI.2017.7950480.

[27] I. Sutskever; J. Martens; G. Dahl;G. Hinton. On the importance of initialization and momentum in deep learning. *Proceedings of the 30th International Conference on Machine Learning*, 2013. http://www.cs.toronto.edu/ fritz/absps/momentum.pdf, last accessed on 29.08.2018.

[28] T. Schlegl; Ph. Seeböck; S. M Waldstein; U. Schmidt-Erfurth; G. Langs. Unsupervised Anomaly Detection with Generative Aversarial Networks to Guide Marker Discovery. *published in proceedings in IPMI 2017*, 2017. https://arxiv.org/pdf/1703.05921.pdf, last accessed 29.08.2018.

[29] A. Collignon; D. Vandermeulen; P. Suetens; G. Marchal. 3d multimodality medical image registration using feature space clustering. *Computer Vision, Virtual Reality, and Robotics in Medicine*, 905:195–204, 1995.

[30] A. Collignon; F. Maes; D. Delaere; D. Vandermeulen; P. Suetens; G. Marchal. Automated multi-modality image registration based on information theory. *Proc.Int. Conf. Inf.Process. Med. Imag.*, pages 263–274, 1995.

[31] C. A. Méndez; F. P. Ferrarese; P. Summers; G. Petralia; M. Bellomi; G. Menegaz. Multimodal MRI-based tissue classification in breast ductal carcinoma. *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 2012. DOI: 10.1109/ISBI.2012.6235504.

[32] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning.* MIT Press, 2016. `http://www.deeplearningbook.org`.

[33] C. K. Kuhl ; H. H. Schild. Dynamic image interpretation of MRI of the breast. *Journal of Magnetic Resonance Imaging*, 12:965–974, 2000.

[34] L. Wang; M. Harz; T. Boehler; B. Platel; A. Homeyer; H. K. Hahn. A robust and extendable framework towards fully automated diagnosis of nonmass lesions in breast DCE-MRI. 2014. DOI: 10.1109/ISBI.2014.6867826.

[35] J. Kim; J. A. Fessler. Intensity-Based Image Registration Using Robust Correlation Coefficients. *IEEE Transactions on Medical Imaging*, 23(11):1430–1444, 2004.

[36] B. B. Avants; N. J. Tustison; G. Song; Ph. A. Cook; A. Klein; J. C. Gee. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuoimage*, 54(3):2033–2044, 2011.

[37] N. J. Tustison; B. B. Avants; Ph. A. Cook; Y. Zheng; A. Egan; P. A. Yushkevich; J. C. Gee. N4ITK:Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010.

[38] D. P. Kingma; J. L. Ba. Adam: A method for stochastic optimization. 2014. arXiv:1412.6980v9 [cs.LG]
https://arxiv.org/abs/1412.6980, last accessed on 19.08.2018.

[39] X. Zhang; Y. Zhang; E. Y. Han; N. Jacobs; Q. Han; X. Wang; J. Liu. Whole Mammogram Image Classification With Convolutional Neural Networks. *International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017. DOI: 10.1109/BIBM.2017.8217738.

[40] K. He; X. Zhang; S. Ren; J. Sun. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. 2015. arXiv:1502.01852v1 [cs.CV], https://arxiv.org/abs/1502.01852, last accessed on 29. 08. 2018.

[41] Sh. C. Agner; J. Xu; H. Fatakdawala; Sh. Ganesan; A. Madabhushi; S. Englander; M. Rosen; K. Thomas; M. Schnall; M. Feldman; J. Tomaszewski. Segmentation and classification of triple negative breast cancers using DCE-MRI . *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009. DOI: 10.1109/ISBI.2009.5193283.

[42] B.B. Avants; C.L. Epstein; M. Grossman ; J.C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12:26–41, 2008.

[43] J.P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2:243–260, 1998.

[44] H. Lee; K. Lee; M. Ko; J. Kang; I. Joo; H. Moon; K.–M. Kim. 3D Breast Registration for PET-CT and MR Based on Surface Matching. *IEEE Nuclear Science Symposium Conference Record*, 2011. DOI: 10.1109/NSSMIC.2011.6152567.

[45] G. J. Wengert; T. H. Helbich; W.–D. Vogl; P. Baltzer; G. Langs; M. Weber; W. Bogner; St. Gruber; S. Trattnig; K. Pinker. Introduction of an Automated

User–Independent Quantitative Volumetric Magnetic Resonance Imaging Breast Density Measurement System Using the Dixon Sequence. *Investigative Radiology*, 50(2), 2015.

[46] M. Arjovsky; S. Chintala; L. Bottou. Wasserstein GAN. 2017. arXiv:1701.07875v3 https://arxiv.org/abs/1701.07875, last accessed on 29.08. 2018.

[47] J. P. W. Pluin; A. Maintz; M. A. Viergever. Mutual-Information-Based registration of Medical Images: A Survey. *IEEE Transactions on medical Imaging*, 22(8):986–1004, 2008.

[48] R. A Yeh; Ch. Chen; T. Y. Lim; A. G. Schwing; M. Hasegawa-Johnson; M. N. Do. Semantic image inpainting with perceptual and contextual losses. 2016. arXiv:1607.07539v3
https://arxiv.org/abs/1607.07539, last accessed on 29.08.2018.

[49] R. C. Conceição; R. M. Capote; B. L. Oliveira; M. Glavin; E. Jones; M. O'Halloran. Imaging and classification of breast cancer with multimodal PEM-UWB techniques. *2013 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, pages 421 – 424, 2013. DOI: 10.1109/ICEAA.2013.6632271.

[50] T. Vercauteren X. Pennec; A. Perchant; N. Ayache. Non-parametric Diffeomorphic Image Registration with the Demons Algorithm. *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, pages 319–326, 2007.

[51] N. Otsu. A thresholf selection method from graylevel histogramms. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, 1979.

[52] A. Sotiras; C. Davatzikosa; N. Paragios. Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013.

[53] M. Wodzinski; A. Skalski ; I. Ciepiela; T. Kuszewski; P. Kedzierawski. Application of Demons Image Registration Algorithms in Resected Breast Cancer Lodge Localization. *2017 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pages 400 – 405, 2017. DOI: 10.23919/SPA.2017.8166900.

[54] K. Kuczynski; M. Siczeky; R. Stegierski; P. Mikolajczak. Application of Image Registration Techniques in Breast Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pages 2167 – 2171, 2011. DOI: 10.1109/FSKD.2011.6019985.

[55] A. Gruslys; J. Acosta-Cabronero; P. J. Nestor G. B. Williams; R. E. Ansorge. A New Fast Accurate Nonlinear Medical Image Registration Program Including Surface Preserving Regularization. *IEEE Transactions on Medical Imaging*, 33(11), 2014. DOI: 10.1109/TMI.2014.2332370.

[56] W. M. Wells III; P. Viola; H. Atsumi; S. Nakajima; R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Med. Image Anal.*, 1(1):35–51, 1996.

[57] B. Platel; R. Mus T. Welte; N. Karssemeijer; R. Mann. Automated Characterization of Breast Lesions Imaged With an Ultrafast DCE-MR Protocol. *IEEE Transactions on Medical Imaging*, 33(2), 2014.

[58] A. Klein; J. Andersson; B. A. Ardekani; J. Ashburner; B. Avants; M.–Ch. Chiang; G. E. Christensen; D. L. Collins; J. Gee; P. Hellier; J. Hyun Song; M. Jenkinson; C. Lepage; D. Rueckert; P. Thompson; T. Vercauteren; R. P. Woods; J. J. Mann R. V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuoimage 64(3)*, 64(3):786–802, 2009.

[59] D. Rumelhart; G. Hinton; R. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[60] A. Radford; L. Metz; S. Chintala. Unsupervised representation learning with deep convolutional generative adversaral networks. 2014. arXiv:1511.06434v2, https://arxiv.org/abs/1511.06434, last accessed on 29.08.2018.

[61] Modat; T. Vercauteren; G. R. Ridgway; D. J. Hawkes; N. C. Fox; S. Ourselin. Diffeomorphic Demons using Normalised Mutual Information, Evaluation on Multi-Modal Brain MR Images. *Proceedings of SPIE - The International Society for Optical Engineering*, 7623, 2010. DOI: 10.1117/12.843962.

[62] Y. Guo; R. Sivaramakrishna; Ch.–Ch. Lu; J. S. Suri; Swamy Laxminarayan. Breast image registration techniques: a survey. *Med. Biol. Eng. Comput.*, 44:15–26, 2006.

[63] H. J. Johnson; M. M. McCormick; L. Ibánez; the Insight Software Consortium. *The ITK Software Guide Book 2: Design and Functionality Fourth Edition Updated for ITK version 4.9* . January 25, 2016. http://cvinhais.noip.me/materials/itk/InsightSoftwareGuide-Book2-4.9.0.pdf, last accessed on 2.6. 2018.

[64] W.-D. Vogl. Automatic Segmentation and Classification Of Breast Lesions Using a Novel Multimodal Imaging Approach. Master thesis at Technical University of Vienna, 2012.

[65] N. V. Ruiter; R. Stotzka; T.–O. Müller; H. Gemmeke; J. R. Reichenbach; W. A. Kaiser. Model-Based Registration of X-Ray Mammograms and MR Images of the Female Breast. *IEEE Transactions on Nuclear Science*, 53(1), 2006.

[66] Valerie Wiesner. Classification of background parenchymal enhancement through quantification and texture analysis. Master thesis at Medical University of Vienna, 2017.

[67] T. Salimans; I. Goodfellow; W. Zaremba; V. Cheung; A. Radford; X. Chen. Improved Techniques for Training GANs. *Advances in Neural Information Processing Systems*, pages 2226–2234, 2016.

[68] M. Abadi; A. Agarwal; P. Barham; E. Brevdo; Z.Chen; C. Citro; G. S. Corrado; A. Davis; J. Dean; M. Devin S. Ghemawat I. Goodfellow A. Harp; G. Irving; M. Isard; Y. Jia; R. Jozefowicz; L. Kaiser; M. Kudlur; J. Levenberg; D. Mané; R. Monga; Sh.Moore; D. Murray; Ch.Olah ; M. Schuster; J. Shlens; B. Steiner; I. Sutskever; K. Talwar; P. Tucker; V. Vanhoucke; V.Vasudevan; F. Viégas; O. Vinyals P. Warden M. Wattenberg; M. Wicke; Y. Yu; X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. paper: http://download.tensorflow.org/paper/whitepaper2015.pdf last accessed on 30.8.2018,
Software available from tensorflow.org: https://www.tensorflow.org/.

[69] I. J. Goodfellow; D. Warde-Farley; M. Mirza; A. Courville; Y. Bengio. Maxout networks. *JMLR WCP*, 8:1319–1327, 2013. https://arxiv.org/pdf/1302.4389.pdf, last accessed on 30.8.2018.

[70] I. J. Goodfellow; J. Pouget-Abadie; M. Mirza; B. Xu; D. Warde-Farley; S. Ozair; A. Courville; Y. Bengio. Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[71] K. Jarrett; K. Kavukcuoglu; M. Ranzato; Y. LeCun. What is the best multi-stage architecture for object recognition? *2009 IEEE 12th International Conference on Computer Vision*, 2009. DOI: 10.1109/ICCV.2009.5459469.