COMPUTER AIDED DETECTION FOR BREAST LESIONS IN ULTRASOUND AND MAMMOGRAPHY

Richa Agarwal

DOCTORAL THESIS

Computer Aided Detection for Breast Lesion in Ultrasound and Mammography

Richa Agarwal

2019

DOCTORAL THESIS

Computer Aided Detection for Breast Lesion in Ultrasound and Mammography

Richa Agarwal

2019

DOCTORAL PROGRAM IN TECHNOLOGY

Supervisors:
Dr. Robert Martí,
Dr. Oliver Díaz and Prof. Xavier Lladó

This thesis is presented in fulfillment of the requirement for the conferral of the degree of Doctor of Philosophy by the University of Girona

Dr. Robert Martí, Dr. Oliver Díaz and Prof. Xavier Lladó from University of Girona

I / WE DECLARE

That the work entitled *Computer Aided Detection for Breast Lesion in Ultrasound and Mammography* presented by *Richa Agarwal* to obtain the degree of Doctor of Philosophy has been developed under our supervision and complies with the requirements needed to obtain the International Mention.

Therefore, in order to certify the aforesaid statements, we sign the document.

Girona, September 2019.

This work is dedicated to my parents and loving husband. . . . . .

Acknowledgments

First and foremost, I want to thank my supervisors, Dr. Robert Martí, Dr. Oliver Díaz and Prof. Xavier Lladó, for giving me an opportunity to work under their guidance and provide continual support throughout my thesis. Their patience, motivation and enthusiasm deserve special gratitude. I am indebted to them for guidance, expertise and for providing me with a conducive atmosphere to carry out my research. It is an humbling experience working with them, to say the least.

I want to express my gratitude to the Generalitat de Catalunya for awarding me with the research grant FI-DGR2016, which has been used to fund this doctoral thesis. The image datasets used in this work have been gently facilitated by different research hospital centres. One of the important experiences during my PhD was my visit to Manchester Metropolitan University, United Kingdom. I would like to give a special thank to Dr. Moi Hoon Yap, my advisor. Her time, dedication and contributions to this thesis are invaluable.

I would like to thank my colleagues and friends who have been part of ViCOROB: Sergi, Mariano, Sandra, Moustafa, Kaiser, Jose, Juan, Dina, Klemen, Khadidja, and many more. I will always remember our long lasting discussion on various topics. I learned so much from all of you. Lastly I would like to thank my family for their unconditional support both before and during my whole career, especially I would like to pay my special regards to my husband Dr. Dheeraj Agarwal, who was always there to help and encourage me.

List of Acronyms

ABUS Automated Breast Ultrasound

AD Anisotropic Diffusion

AD-LBR Anisotropic Diffusion-Lattice based Reduction

ANN Artificial Neural Network

API Application Programming Interface

AUROC Area Under Receiver Operating Curve

BI-RADS Breast Imaging Reporting and Data System

CAD Computer Aided Detection

CC Cranio-Caudal

CED Coherence Enhancing Diffusion

CNN Convolutional Neural Network

DBT Digital Breast Tomosynthesis

DSC Dice Similarity Coefficient

EED Edge Enhancing Diffusion

FC Fully Connected

FFDM Full-Field Digital Mammogram

FN False Negatives

FP False Positives

FPI False Positives per Image

FROC Free-Response Operating Curve

GCN Global Contrast Normalisation

GT Ground-Truth

HHUS Hand-Held Ultrasound

IoU Intersection over Union

ITK Insight ToolKit

LBR Lattice Basis Reduction

MLO Medio-Lateral Oblique

MPM Mass Probability Map

MRI Magnetic Resonance Imaging

OMI-DB OPTIMAM Mammography Image Database

R-CNN Region-Based CNN

ROC Receiver Operating Curve

ROI Region of Interest

RPN Region Proposal Network

SFM Screen-Film Mammogram

TN True Negatives

TP True Positives

TPR True Positive Rate

US Ultrasound

WAT Watershed

List of Tables

List of Figures

Author's Publications

Here we list the international journals and conferences where this work has been published:

Journals

[1] R. Agarwal, O. Diaz, M. H. Yap, X. Lladó, and R. Martí, "Lesion Detection in Whole Mammogram using Faster R-CNN", *IEEE Journal of Biomedical and Health Informatics*, 2019 (Under Review).

[2] R. Agarwal, O. Diaz, X. Lladó, M. H. Yap, and R. Martí, "Automatic mass detection in mammograms using deep convolutional neural networks", *Journal of Medical Imaging*, vol. 6, no. 3, 2019. DOI: 10.1117/1.JMI.6.3.031409.

[3] R. Agarwal, O. Diaz, X. Lladó, A. Gubern-Mérida, J. C. Vilanova, and R. Martí, "Lesion Segmentation in Automated 3D Breast Ultrasound: Volumetric Analysis", *Ultrasonic imaging*, vol. 40, no. 2, pp. 97–112, 2018. DOI: 10.1177/0161734617737733.

Conferences

[4] R. Agarwal, O. Diaz, X. Lladó, and R. Martí, "Mass detection in mammograms using pre-trained deep learning models", in *14th International Workshop on Breast Imaging (IWBI 2018), awarded with travel grant*, International Society for Optics and Photonics, vol. 10718, Atlanta, Georgia, 2018. DOI: 10.1117/12.2317681.

[5] R. Agarwal, O. Diaz, X. Lladó, A. Gubern-Mérida, J. C. Vilanova, and R. Martí, *Segmentación semiautomática en ecografía mamaria automatizada (ABUS)*, Poster at the Congreso Nacional Sociedad Española de Radiología Médica (SERAM), Pamplona, Spain, May 2018.

[6] R. Agarwal, O. Diaz, Y. Díez, A. Gubern-Mérida, J. C. Vilanova, and R. Martí, *A clinical tool for temporal analysis in 3d automated breast ultrasound*, Presentation at the European Congress of Radiology (ECR) selected in 32 best student abstract with the travel grant, Vienna, Austria, Mar. 2017.

[7] O. Diaz, R. Agarwal, A. M. Gubern, J. V. Zelst, Y. Díez, and R. Martí, *Automated volumetric lesion quantification in automated 3D breast ultrasound: Comparison of 5 breast lesion segmentation algorithms*, Presentation and abstract at the European Congress of Radiology (ECR), Vienna, Austria, 2016.

Contents

Abstract

In the field of breast cancer imaging, traditional Computer Aided Detection (CAD) systems were designed using limited computing resources and used scanned films (poor image quality), resulting in less robust application process. Currently, with the advancements in technologies, it is possible to perform 3D imaging and also acquire high quality Full-Field Digital Mammogram (FFDM).

Automated Breast Ultrasound (ABUS) has been proposed to produce a full 3D scan of the breast automatically with reduced operator dependency. When using ABUS, lesion segmentation and tracking changes over time are challenging tasks, as the 3D nature of the images make the analysis difficult and tedious for radiologists. One of the goals of this thesis is to develop a framework for breast lesion segmentation in ABUS volumes. The 3D lesion volume in combination with texture and contour analysis, could provide valuable information to assist radiologists in the diagnosis.

Although ABUS volumes are of great interest, x-ray mammography is still the gold standard imaging modality used for breast cancer screening due to its fast acquisition and cost-effectiveness. Moreover, with the advent of deep learning methods based on Convolutional Neural Network (CNN), the modern CAD systems are able to learn automatically which imaging features are more relevant to perform a diagnosis, boosting the usefulness of these systems. One of the limitations of CNNs is that they require large training datasets, which are very limited in the field of medical imaging.

In this thesis, the issue of limited amount of dataset is addressed using two strategies: (i) by using image patches as inputs rather than full sized image, and (ii) use the concept of transfer learning, in which the knowledge obtained by training for one task is used for another related task (also known as domain adaptation). In this regard, firstly the CNN trained on a very large dataset of natural images is adapted to classify between mass and non-mass image patches in the Screen-Film Mammogram (SFM), and secondly the newly trained CNN model

is adapted to detect masses in FFDM. The prospects of using transfer learning between natural images and FFDM is also investigated. Two public datasets CBIS-DDSM and INbreast have been used for the purpose. In the final phase of research, a fully automatic mass detection framework is proposed which uses the whole mammogram as the input (instead of image patches) and provides the localisation of the lesion within this mammogram as the output. For this purpose, OPTIMAM Mammography Image Database (OMI-DB) is used.

The results obtained as part of this thesis showed higher performances compared to state-of-the-art methods, indicating that the proposed methods and frameworks have the potential to be implemented within an advanced CAD systems, which can be used by radiologists in the breast cancer screening.

Resumen

En el campo de las imágenes de cáncer de mama, los sistemas tradicionales de detección asistido por ordenador (del inglés CAD) se diseñaron utilizando recursos informáticos limitados y películas de mamografía escaneadas (del inglés SFM) de calidad de imagen deficiente, lo que dió como resultado aplicaciones poco robustas. Actualmente, con los avances de las tecnologías, es posible realizar imágenes médicas en 3D y adquirir mamografía digital (del inglés FFDM) de alta calidad.

El ultrasonido automático de la mama (del inglés ABUS) ha sido propuesto para adquirir imágenes 3D de la mama con escasa dependencia del operador. Cuando se usa ABUS, la segmentación y seguimiento de lesiones en el tiempo son tareas complicadas, ya que la naturaleza 3D de las imágenes hace que el análisis sea difícil y tedioso para los radiólogos. Uno de los objetivos de esta tesis es desarrollar un marco para la segmentación semi-automática de lesiones mamarias en volúmenes ABUS. El volumen de la lesión 3D, en combinación con el análisis de la textura y el contorno, podría proporcionar información valiosa para realizar el diagnóstico radiológico.

Aunque los volúmenes de ABUS son de gran interés, la mamografía de rayos x sigue siendo la modalidad de imagen estándar utilizada para la detección precoz del cáncer de mama, debido principalmente a su rápida adquisición y rentabilidad. Además, con la llegada de los métodos de aprendizaje profundo basados en redes neuronales convolucionales (del inglés CNN), los sistemas modernos CAD pueden aprender automáticamente que características de la imagen son más relevantes para realizar un diagnóstico, lo que aumenta la utilidad de estos sistemas. Una de las limitaciones de la CNN es que requiere de grandes conjuntos de datos para entrenamiento, los cuales son muy limitados en el campo de la imagen médica.

En esta tesis, el tema de la poca disponibilidad de imágenes médicas se aborda mediante dos estrategias: (i) utilizando trozos de imagen como entradas en lugar de imágenes de tamaño original, y (ii) mediante técnicas de aprendizaje por transferencia, en el que el conocimiento obtenido mediante la capacitación para una

tarea se transfiere a otra tarea relacionada (también conocido como adaptación de dominio). En primer lugar la CNN entrenada en un conjunto de datos muy grandes de imágenes naturales es adaptada para clasificar entre trozos de imagen de tumores y no tumores en SFM, y en segundo lugar, la CNN entrenada es adaptada para detectar tumores en FFDM. También se investigó el aprendizaje por transferencia entre imágenes naturales y FFDM. Se han utilizado dos conjuntos de datos públicos (CBIS-DDSM e INbreast) para este propósito. En la fase final de la investigación, se propone un marco de detección automática de tumores utilizando la mamografía original como entrada (en lugar de trozos de imagen) y proporciona la localización de la lesión dentro de esta mamografía como salida. Para este propósito, se utiliza otra base de datos (OMI-DB).

Los resultados obtenidos como parte de esta tesis mostraron mejores rendimientos en comparación con el estado del arte, lo que indica que los métodos y marcos propuestos tienen el potencial de ser implementados dentro de los sistemas CAD avanzados, que pueden ser utilizados por radiólogos en el cribado del cáncer de mama.

Resum

En el camp de les imatges de càncer de mama, els sistemes tradicionals de detecció assistida per ordinador (del anglès CAD) es van dissenyar utilitzant recursos informàtics limitats i pel·lícules de mamografia escanejades (del anglès SFM) de qualitat d'imatge deficient, fet que va resultar en aplicacions poc robustes. Actualment, amb els avanços de les tecnologies, és possible realitzar imatges mèdiques en 3D i adquirir mamografies digitals (de l'anglès FFDM) d'alta qualitat.

L'ultrasò automàtic de la mama (del anglès ABUS) ha estat proposat per adquirir imatges 3D de la mama amb escassa dependència del operador. Quan s'utilitza ABUS, la segmentació i seguiment de les lesions en el temps són tasques complicades ja que la naturalesa 3D de les imatges fa que l'anàlisi sigui difícil i feixuc per els radiòlegs. Un dels objectius d'aquesta tesi és desenvolupar un marc per la segmentació semi-automàtica de lesions mamàries en volums ABUS. El volum de lesió 3D, en combinació amb l'anàlisi de la textura i el contorn, podria proporcionar informació valuosa per realitzar el diagnòstic radiològic.

Tot i que els volums de ABUS són de gran interès, la mamografia de raigs X continua essent la modalitat d'imatge estàndard utilitzada per la detecció precoç del càncer de mama, degut principalment a la seva ràpida adquisició i rentabilitat. A més, amb l'arribada dels mètodes d'aprenentatge profund basats en xarxes neuronals convolucionals (del anglès CNN), els sistemes CAD moderns poden aprendre automàticament quines característiques de la imatge són més rellevants per realitzar un diagnòstic, fet que augmenta la utilitat d'aquests sistemes. Una de les limitacions de les CNN es que requereixen de grans conjunts de dades per entrenar, els quals són molt limitats en el camp de la imatge mèdica.

En aquesta tesi, el tema de la poca disponibilitat d'imatges mèdiques s'aborda mitjançant dues estratègies: (i) utilitzant regions de la imatge com a entrada en comptes de les imatges de mida original, i (ii) mitjançant tècniques d'aprenentatge per transferència, en el que el coneixement après per a una determinada tasca es transfereix a una altra tasca relacionada (també conegut com a adaptació de

domini). En primer lloc la CNN entrenada en un conjunt de dades molt gran d'imatges naturals és adaptada per classificar regions de la imatge en tumor i no tumor de SFM i, en segon lloc, la CNN entrenada és adaptada per detectar tumors en FFDM. També s'ha investigat l'aprenentatge per transferència entre imatges naturals i FFDM. S'han utilitzat dos conjunts de dades públiques (CBIS-DDSM i INbreast) per aquest propòsit. En la fase final de la investigació, es proposa un marc de detecció automàtica de tumors utilitzant la mamografia original com entrada (en lloc de regions de la imatge) i que proporciona la localització de la lesió dins d'aquesta mamografia com a sortida. Per aquest propòsit s'utilitza una altra base de dades (OMI-DB).

Els resultats obtinguts com a part d'aquesta tesi mostren millors rendiments en comparació amb l'estat de l'art, el que indica que els mètodes i marcs proposats tenen el potencial de ser implementats dins de sistemes CAD avançats, que poden ser utilitzats per radiòlegs en el cribatge del càncer de mama.

Chapter 1

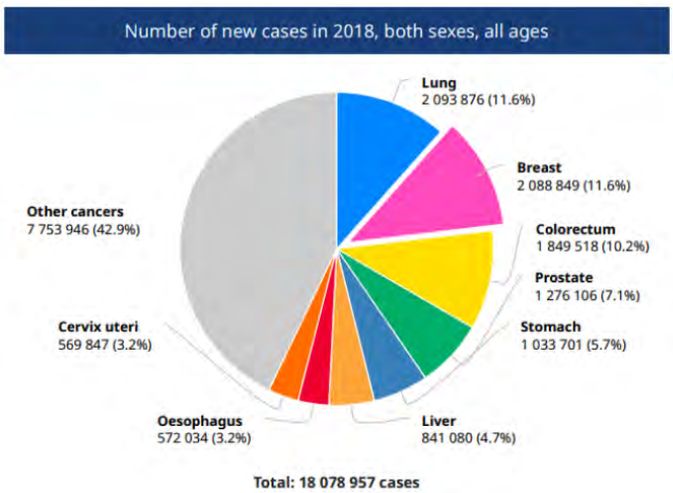
Introduction

The aim of this PhD thesis is the development of Computer Aided Detection (CAD) tools for breast lesions in ultrasound and mammography images. This initial chapter is an introduction to breast cancer, and the different imaging techniques for detecting it. An introductory explanation about the breast abnormalities and CAD systems is also provided, followed by the aims and objectives of this thesis.

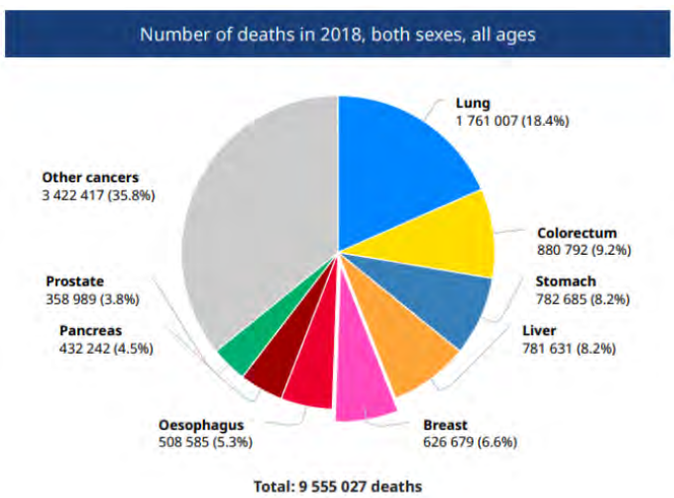
1.1 Breast Cancer

Breast cancer is the most common form of cancer in the female population. In the United State of America, it is estimated that approximately 12% of women will be diagnosed with breast cancer at some point during their lifetime [1]. According to other studies, breast cancer has the highest incidence and death rates among all other cancers (excluding Melanoma skin cancer) [2]. An overview of breast cancer cases in the world is shown in Fig. 1.1.

In the European Union (EU), breast cancer is also the leading cause of death among the female population. In 2015, deaths from breast cancer made up around 7.2% of all deaths from cancer; among women breast cancer accounted for 15.6% of all deaths from cancer (see Fig. 1.2) [4]. The standardised death rate from breast cancer in EU was 32.7 deaths per 100,000 female inhabitants. Among the EU Member States, the highest standardised death rate for breast cancer among women was recorded in Croatia (43.1 per 100,000 inhabitants), followed by Slovakia and



(a)



(b)

Figure 1.1: Estimate of breast cancer cases in the World, (a) incidence, and (b) mortality [3].

Hungary, just surpassing 40 per 100,000 inhabitants in the former while remaining slightly below this level in the latter. Five EU member states recorded standardised death rates for breast cancer that were below 30 per 100,000 inhabitants: the Czech Republic, Sweden, Portugal and Finland, with the lowest rate recorded in Spain (23.4 per 100,000 inhabitants) [5].

Breast cancer incidence has increased in the past decade. This can be attributed to the introduction of screening programs which results in an early detection of cancers. Extensive efforts are being made to reduce the mortality rate due to breast cancer. It is a known fact that the survival rate is dependent on the stage at which cancer is diagnosed, therefore it is necessary to detect the cancer in early stage rather than detecting at the later stages when it becomes difficult to cure.

	Number of deaths	Standardised death rates				
		Total	Males	Females	Persons aged < 65 years	Persons aged 65 and over
	(number)	(per 100 000 inhabitants)				
EU-28	95 265	18.8	0.5	32.7	7.0	67.4
Belgium	2 196	20.0	0.4	35.2	7.3	72.2
Bulgaria	1 343	18.7	1.0	31.9	7.8	63.7
Czech Republic	1 623	17.3	0.4	29.0	5.4	66.5
Denmark (1)	1 069	20.4	0.2	36.6	5.9	80.6
Germany	18 320	20.6	0.4	35.9	6.9	77.2
Estonia	245	19.4	0.6	30.2	7.9	66.6
Ireland	678	20.2	0.1	37.0	7.8	71.8
Greece	2 135	17.9	0.4	32.0	6.2	66.2
Spain	6 293	13.2	0.4	23.4	5.6	44.8
France	12 611	19.0	0.6	33.0	7.3	67.2
Croatia	1 054	26.4	0.9	43.1	8.2	101.3
Italy	12 390	18.0	0.4	31.5	6.8	63.8
Cyprus	113	16.8	0.3	31.0	6.4	59.8
Latvia	447	22.3	0.3	35.2	9.5	75.4
Lithuania	571	19.8	0.4	31.6	8.9	64.7
Luxembourg (1)	93	20.4	0.5	37.1	7.7	72.5
Hungary	2 248	24.1	0.8	39.3	9.0	86.8
Malta	81	20.5	:	36.1	6.6	77.7
Netherlands	3 311	21.1	0.4	37.7	8.0	75.0
Austria	1 595	19.0	0.7	32.5	6.3	71.7
Poland	6 394	19.3	0.5	32.1	7.6	67.7
Portugal	1 704	15.7	0.4	26.8	6.6	53.4
Romania	3 504	19.0	0.8	32.7	8.4	62.6
Slovenia	433	21.7	0.2	36.6	6.9	82.8
Slovakia	1 040	24.8	0.6	40.6	8.0	94.3
Finland	823	14.8	0.3	26.1	5.3	54.2
Sweden	1 430	14.8	0.2	27.0	5.1	55.1
United Kingdom	11 521	19.1	0.3	34.4	7.3	68.1
Iceland (2)	54	21.2	1.1	40.0	8.5	73.7
Liechtenstein	6	17.3	:	32.6	9.8	48.1
Norway	595	13.2	0.4	23.8	4.9	47.8
Switzerland	1 433	18.3	0.2	32.5	5.4	71.7
Serbia	1 729	24.6	1.0	43.0	10.5	82.9
Turkey	3 914	9.0	0.6	15.9	4.6	27.2

(1) Males: 2014.
(2) Males: 2013.
Source: Eurostat (online data codes: hltl

Figure 1.2: Breast cancer mortality statistics from European Union [4].

According to European Breast Cancer guidelines, it is highly recommended that women between 50 and 69 years old have mammography screening for breast cancer [6]. The risk of dying from breast cancer is reduced by between 10 (low risk population) and 50 (high risk population) per 10,000 women offered screening. This corresponds to a reduction of 10 to 60 breast cancer deaths per 10,000 in women actually screened.

1.2 Breast Anatomy

The beast is a highly complex structure, consisting of around 15 to 20 sections called lobes. Each lobe consists of many smaller structures called lobules. The lobules are arranged in clusters, and at the end of each lobule, there are tiny “bulbs” that produce milk. The lobes, lobules and bulbs are connected together by small tubes called ducts. The ducts carry milk to the nipples, which is located in the

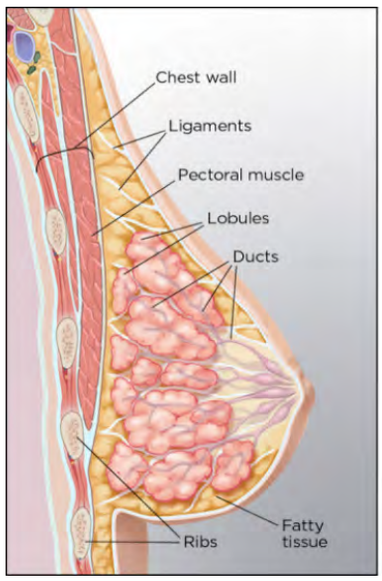


Figure 1.3: Illustration showing the breast anatomy [7].

middle of the areola (the darker area surrounding the nipple). The spaces between the lobes and ducts are filled with fats. In Fig. 1.3 the breast anatomy is illustrated depicting the inside structure of the breast. The female breasts contain different types of fatty, fibrous, and glandular tissue:

- glandular tissue includes the breast lobes and breast ducts.
- fibrous tissues include ligaments, supportive tissues (dense breast tissue) and scar tissues.
- fatty tissue (non-dense breast tissue) fills in the spaces between glandular and fibrous tissue and largely determines the breast size.

In general, all non-fatty tissue are called as fibroglandular tissue. In addition, there are also bands of supportive, flexible connective tissue called ligaments, which stretch from the skin to the chest wall to hold the breast tissue in place. The pectoral muscle lies against the chest wall underneath both breasts, giving them support. Moreover, no two women's breasts are the same as each contain a specific combination of fatty and dense tissue. Some women's breasts are almost all fatty, whereas others have varying proportions of fatty and fibroglandular tissue. As the age of women increases, the proportion of fatty tissue gradually increases, so that by the age of 70 approximately 80% of all women have breasts that are composed of mostly fatty tissues. The cancer can form in any part of the breast, and appropriate imaging techniques are required for breast cancer screening in women.

1.3 Breast Cancer Imaging Techniques

The conventional imaging modality used for breast cancer screening is x-ray mammography, as it is a fast and cost effective way for screening large population. The diagnosis process becomes more difficult for women with dense breasts, who are at a higher risk (4-6 times) of having cancerous tumour than a women with fatty or non-dense breast [8]. Since, traditional mammography is a 2D imaging technique, the tissue superposition in the dense breasts leads to a poor diagnosis performance [9, 10]. A possible solution to improve cancer detection in dense breasts is so called, personalized breast cancer screening where different breast imaging technologies are recommended to each individual based on its breast cancer risk [11]. Potential 3D imaging technologies include Automated Breast Ultrasound (ABUS), Digital Breast Tomosynthesis (DBT) and Magnetic Resonance Imaging (MRI).

1.3.1 Mammography

Mammography is a modality used for detecting breast cancer at an early stage using low dose x-ray to project the inner tissues of the breast. This way, signs of malignancy (masses, micro-calcifications, asymmetries and distortions) could be visualised. A sample diagram of a mammography system is shown in Fig. 1.4, where the breast is placed between the compression plate and breast support, and then exposed to low intensity x-ray beams from the top (x-ray tube). An image receptor is placed under the breast support to capture the x-ray photons generated in the tube. In addition, an anti-scatter grid is typically located between the breast support and the detector to reduce the scattered radiation signal, which reduces the contrast in the image. The scanners produces raw (“for processing”) and processed (“for presentation”) mammograms as shown in Fig. 1.5, and in this thesis “for presentation” images are used.

The mammography-based breast cancer screening is the most commonly used technique as it is fast, less expensive and do not require a highly skilled operator. However, there are a number of limitations linked with this technique. In certain female population groups (with dense breasts), it has a high rate of False Negatives (FN: detection shows that there is no cancer but there is a cancer) and False Positives (FP: detection shows some symptoms of cancer (tumour or micro-calcifications) but there is no cancer). In the later, a women with no disease is biopsied, whereas in the former a women with disease is untreated and can lead to a casualty.

In mammography-based screening programmes, each breast is acquired using

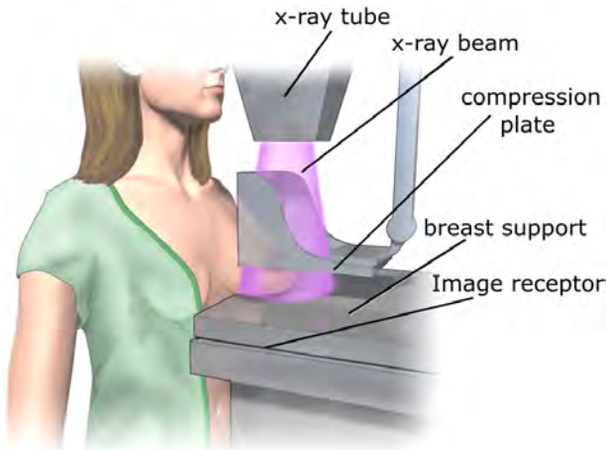


Figure 1.4: An overview of mammography setup, where a women's breast is placed on the breast support and a x-ray beam projection is generated in the x-ray tube. (courtesy: Wikipedia).

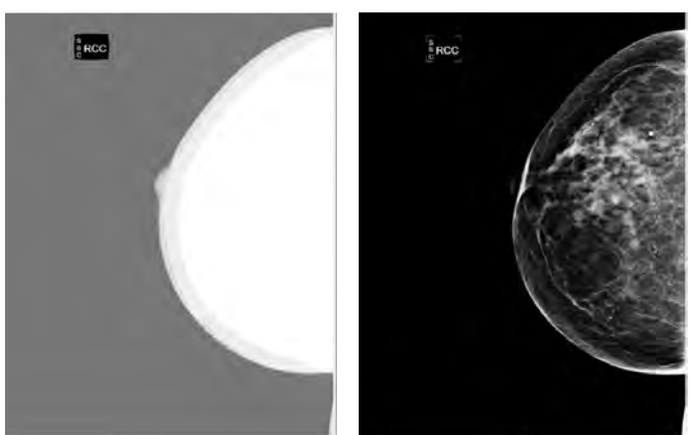


Figure 1.5: Illustration showing the raw and processed mammogram in FFDM.

two different view points (CC and MLO) as shown in Fig. 1.6a. In Fig. 1.6b and Fig. 1.6c the sample mammograms from these two view points are shown. Additional image projections (e.g. magnified views, medio-lateral views) can be performed if the radiologist has observed suspicious regions. Traditionally, Screen-Film Mammogram (SFM) used photographic films to capture the scan of the breast. With the advancements in imaging techniques, currently the high quality Full-Field Digital Mammogram (FFDM) are used which can be directly visualised using computers.

To overcome the problems associated with traditional 2D mammography, pseudo-3D DBT has rapidly developed in the last few years as a new imaging tool to reduce the masking effect of overlapping fibro-glandular tissue, thereby improving breast cancer detection [14–16]. In DBT, images are acquired as the x-ray tube travels across a limited arc above the breast (Fig. 1.7) and multiple low-dose x-ray

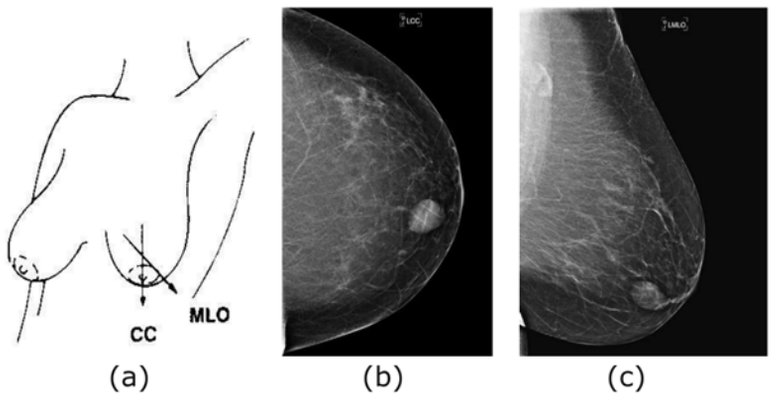


Figure 1.6: Mammography projection views used in breast cancer screening studies: (a) shows the direction of two mostly used view points to produce mammograms, (b) CC view, and (c) MLO view [12].

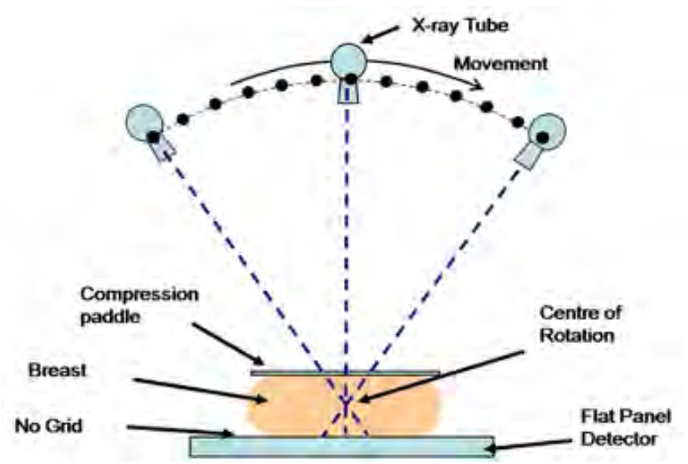


Figure 1.7: Schematic procedure of DBT, showing the movement of the x-ray tube to acquire images at different angles [13]. Note that the geometry differs between manufacturers.

exposures are obtained, which are then post-processed to create pseudo-3D volumes of the breasts. Each slice of such pseudo-3D volume represent a depth of the breast. Slices are typically separated 1 mm between each other, thus producing a better visualisation of the internal breast tissue, and reduces the overlapping effects observed in mammography. Different DBT geometries and acquisition parameters (e.g. narrow vs wide angle) are available, each of which has different detection performance in masses and micro calcifications [17]. Fig. 1.8 shows an example of breast DBT volume with different slices.

Although DBT is gradually being adopted, x-ray mammography is still the gold standard imaging modality used for breast cancer screening due to its fast acquisition and cost-effectiveness. Moreover, DBT is relatively new and its availability at the hospitals is limited. So, it is not yet considered to be a standard method for breast cancer screening.

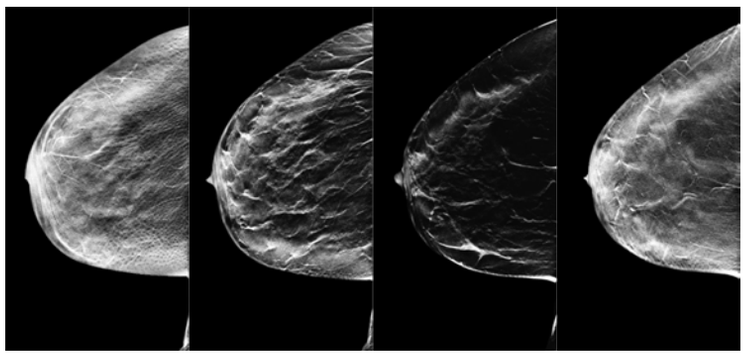


Figure 1.8: Sample DBT volume: Different slices of DBT volume traversing from left to right.

(a) **(b)**

Figure 1.9: (a) A hand-held probe (transducer) used for ultrasound [18], (b) Ultrasound visualization system (source: *google*)

.

1.3.2 Ultrasound Imaging

Ultrasound Imaging or Sonography can be used as an additional modality for the screening examinations. Breast ultrasound imaging uses sound waves to produce a visualization of the breast structure and have been very effective for cancer detection. Since Ultrasound (US) examinations are non-invasive and do not use ionizing radiations (as used in x-ray mammography), it can be considered as a safe procedure. Moreover, US imaging allows operator/radiologist to observe the structure and movement of the body's internal organs in real time.

Conventional US or Hand-Held Ultrasound (HHUS) is performed by a skilled operator using a hand-held probe called transducer (Fig. 1.9a) on top of the skin with aid of a gel (gel reduces the air interference between the transducer and skin and act as a conductive medium for sound waves). High-frequency sound waves are then transmitted through probe to the body. The transducer collects the sound that bounce back, which is then used to create an image on the visualization unit (Fig. 1.9b). Fig. 1.10 shows a sample image obtained from HHUS.

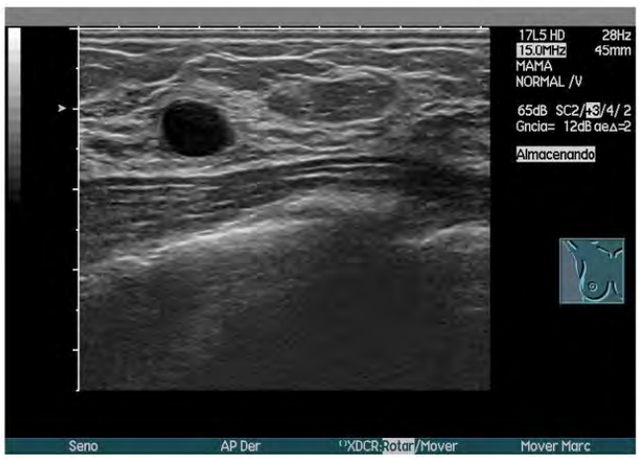


Figure 1.10: Sample breast ultrasound scan [19].

(a) **(b)**

Figure 1.11: (a) ABUS acquisition set-up, (b) Conventional HHUS.

Automated Breast Ultrasound (ABUS) is the latest technology proposed for breast screening and is getting popular especially in the population with the dense breast. ABUS produces a full scan of breast tissues as 3D volume compared to 2D images obtained using HHUS. Currently, there is a lot of focus and development in ABUS technology both clinically as well as in the field of CAD. ABUS follows a process in which the ultrasound gel is applied to the breast and then the scanner, with a much larger transducer (Fig. 1.11a), is placed on the breast to scan the whole breast in slices. Later all the slices are combined to generate a 3D volume of the breast. The transducer is in contact with the skin through a thin membrane which adapts itself with the shape of the breast to capture a full 3D view, overcoming the limitation of HHUS *i.e.* scanning only the suspected area (Fig. 1.11b).

ABUS volumes are visualised using anatomical planes which transect the breast (for reference human body is used in Fig. 1.12a) using three basic reference planes: the Sagittal plane, the Coronal plane, and the Axial/Transverse plane as shown in Fig. 1.12(a). A volumetric visualisation of a ABUS volume in these three planes are shown in Fig. 1.12(b-d).

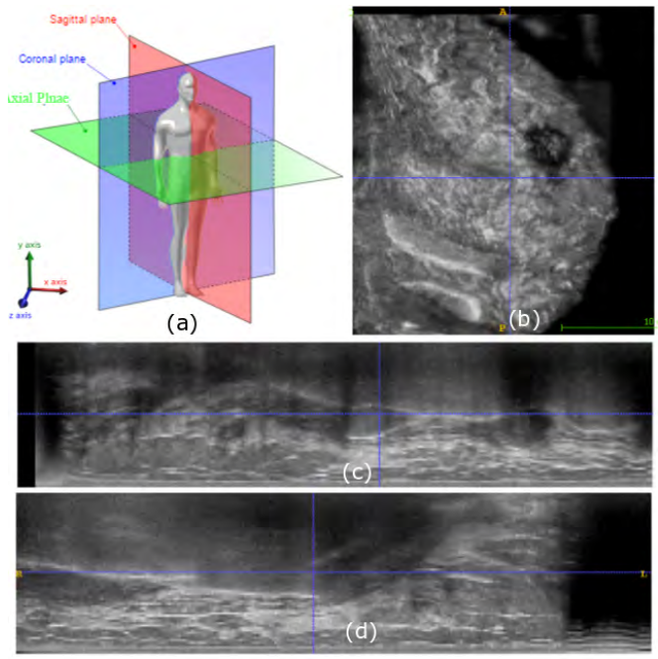


Figure 1.12: Visualization of ABUS volume in three different planes, (a) Anatomical plane depicting 3 different planes: (b) Coronal plane view, (c) Sagittal plane view and (d) Axial plane view of a 3D ABUS volume.

1.3.3 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a radiological technique in which magnetic field and radio waves are used to generate detailed images of the inside of the body. An MRI scan can be used to examine almost any part of the body including heart, brain, bones, spinal cords, breasts etc. Breast MRI is primarily used as a supplemental tool to breast screening with mammography or ultrasound. It may be used to screen women at high risk for breast cancer, evaluate the extent of cancer following diagnosis, or further evaluate abnormalities seen on mammography. Typical breast MRI is performed on a 1.5 Tesla magnet with a dedicated multichannel breast coil. In breast imaging, dynamic contrast-enhanced MRI (DCE-MRI) is used, and is a noninvasive process. Contrast agents play a crucial role in DCE-MRI and should be carefully selected in order to improve accuracy in DCE-MRI examination [20]. There are different image acquisition protocols (sequences) used in breast imaging. The most common ones are T1-, T2-weighted, each of which has been obtained with different acquisition parameters (relaxation time (TR), echo time (TE)).

During the breast MRI scan, the women lies down in prone position on a narrow, flat table such that the breasts hang down into an opening in the table so they can be scanned without being compressed. The table then slides into a

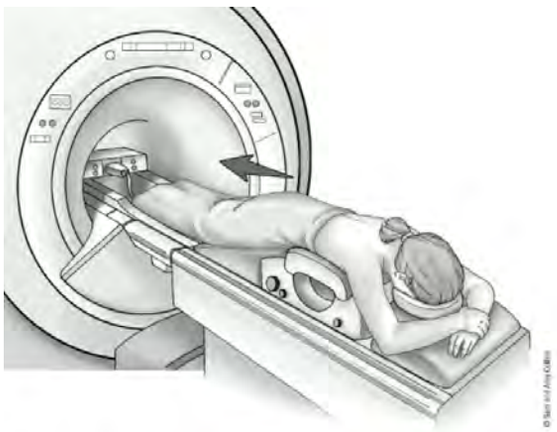


Figure 1.13: Schematic of Breast MRI scan performed, where the patient is lying in prone position.

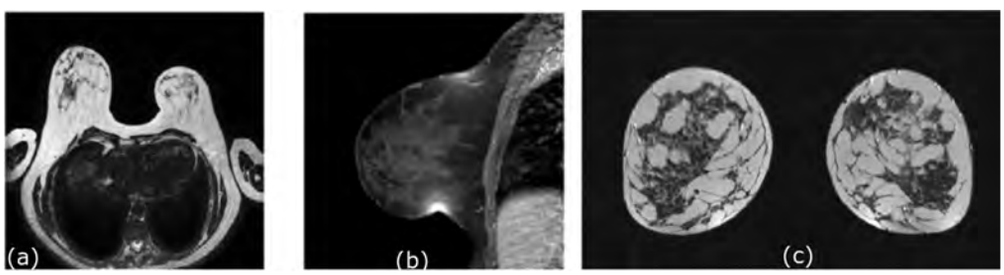


Figure 1.14: Sample of MRI volume showing (a) Axial, (b) Sagittal, and (c) Coronal views [19].

long, narrow cylinder opening of the machine to scan the specific body parts. The machine generates strong magnetic field around the person and radio waves are directed at the body to obtain the scan. A typical breast MRI scanner is shown in Fig. 1.13, and sample of 3D MRI volume with different views are shown in Fig. 1.14.

MRI is not recommended for screening general population because it can miss some cancers that a mammogram would find. Moreover, MRI is a expensive process and is used for young women who are at higher risk of developing cancer. They might be at higher risk due to having family members with cancer or because they have certain gene abnormalities.

1.4 Breast Abnormalities

When a radiologist interprets a mammogram in terms of abnormalities, he or she assigns a score to it which is used to communicate with doctors about how concerned he or she is about the findings. These informations are summed up in one number, called as Breast Imaging Reporting and Data System (BI-RADS)

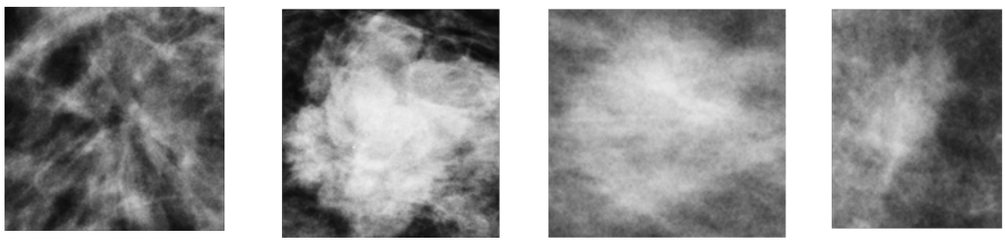


Figure 1.15: Examples of masses in mammograms from OMI-DB dataset [22]

score [21]. The BI-RADS score range from 0 to 6. The category 0 or BI-RADS 0 is utilized when further imaging evaluation (e.g. additional views or ultrasound) or retrieval of prior examinations is required. BI-RADS 1 signifies that there are no abnormalities in the breast, while BI-RADS 2, 3 signify that the findings are considered to be benign. BI-RADS 4 shows suspicious malignancy and BI-RADS 5, 6 signify confirmed malignancy.

In addition, according to BI-RADS there are four other important findings in mammograms: masses or lesions, calcifications, architectural distortions and asymmetries. The architectural distortions and asymmetries are similar in appearance to masses, and can often be confused by the radiologists to be masses.

Masses or Lesions

The mass is a lump that develops in the breast. Most of the masses are non-cancerous (or benign), while some of them could be cancerous in nature (or malignant) and can have life threatening effect if left untreated. In mammography screening, the lesions are usually analysed on two different projections (CC and MLO) and have completely or partially convex-outward borders and radiologically appears denser in the centre than at the periphery. Some examples of masses in a mammogram are shown in Fig. 1.15. If a potential mass is seen only on a single projection, it could be called an asymmetry. Masses can be classified by shape, margin or density.

Calcifications

Breast calcifications are calcium deposits within the breast that appear in mammograms as white spots similar to grains of salt. There are two types of breast calcifications according to their size: macro-calcifications and micro-calcifications. Macro-calcifications look like large white dots or dashes and are almost always non-cancerous. On the contrary, micro-calcifications are very fine white specks and their irregular clustering is usually a sign of cancer. Some examples of calcifications in mammograms are shown in Fig. 1.16.

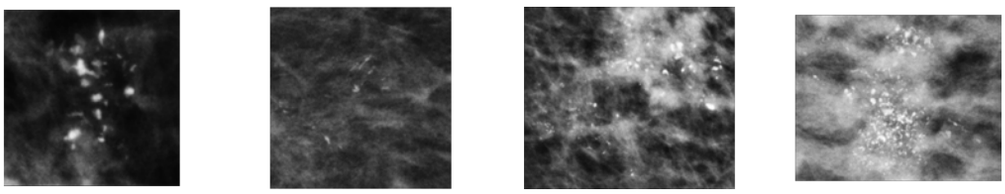


Figure 1.16: Examples of calcifications in mammograms from OMI-DB dataset [22]

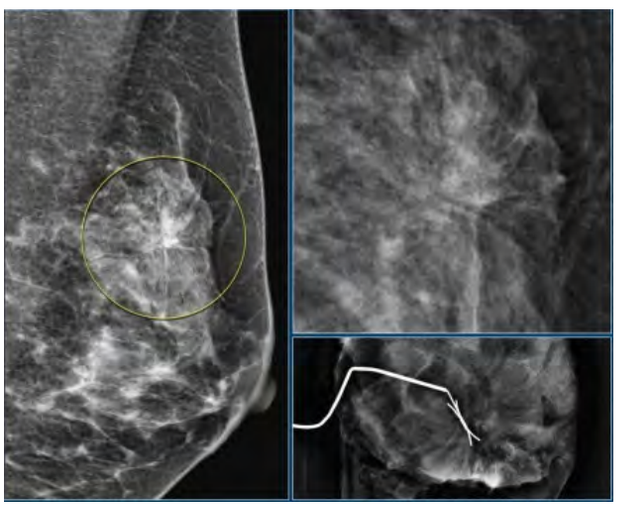


Figure 1.17: Examples of architectural distortion in a mammogram [23]

Architectural Distortions

Architectural distortions are characterised by the distortion of the parenchyma with no definite mass visible. For mammography, this includes thin straight lines or spiculations radiating from a point, and focal retraction, distortion, or straightening at the anterior or posterior edge of the parenchyma. Architectural distortions may also be associated with asymmetry or calcifications. In the absence of an appropriate history of trauma or surgery, architectural distortions are suspicious for malignancy. An example of architectural distortion in the mammogram is shown in Fig. 1.17.

Asymmetries

The internal structure of the two breasts are very similar, so the fact of detecting an asymmetry of the breast parenchyma between left and right breasts may be indicative of the presence of a lesion or the development of a cancer. Depending on the type of asymmetry, this is visible on only one mammographic projection or more than one projections. Focal asymmetry is visible on two projections (Fig. 1.18), while global asymmetry consists of an asymmetry over at least one quarter of the breast (Fig. 1.19).

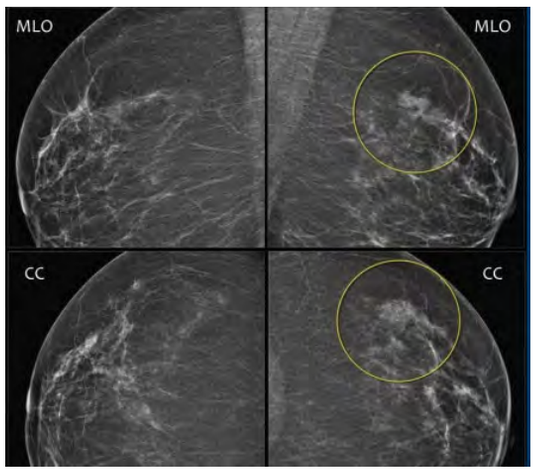


Figure 1.18: Examples of focal asymmetry in a mammogram [23]

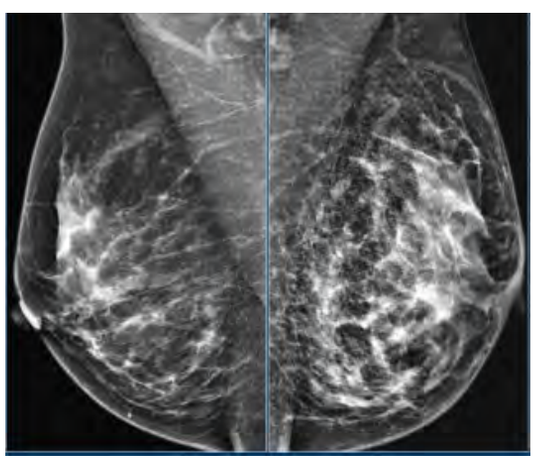


Figure 1.19: Examples of global asymmetry in a mammogram [23]

1.5 Computer Aided Detection

A Computer Aided Detection (CAD) system can be described as the use of computers to evaluate the medical images in an automatic or semi-automatic manner. In the field of breast imaging, availability of accurate CAD methods can make a real impact in improving the current breast screening procedures. Reading and understanding breast images requires a well trained and experienced radiologist, and the CAD systems can be effectively used as the "second opinion" and assist radiologists in screening programs [24, 25].

CAD employs computer vision techniques and/or artificial intelligence to deal with radiological and pathology images. In recent years, with the advancement in technology there has been an increase in the use of different imaging modalities (section 1.3) for detection and diagnosis of cancers/tumours. Reading and understanding the images from these different modalities require highly

experienced and well-trained doctors/radiologists. Moreover, even with well trained experts, there exists high possibilities of intra- and inter-variabilities between the readers [26, 27]. This motivated the use of computers to support radiologists to make accurate diagnosis. The advantage of using a CAD systems is that it can reduce the diagnostic time, reduce inter-observer variations, and can act as a supplementary tool for the radiologists [28–30].

In recent years, with the advancements in computer technology and data sciences, there has been a lot of interest in exploring deep learning methods for various tasks. In this regard, deep learning methods based on Convolutional Neural Network (CNN) have also gained importance in the field of medical image analysis and the efforts are laid to develop modern CAD systems based on the these newly developed CNN algorithms [31, 32]. This thesis is also a effort in this direction, exploring recent advancements in CNN to facilitate the development of an automated CAD system to assist radiologist in fast and accurate detection of lesions during breast cancer screening.

1.6 Thesis Aims and Objectives

The research presented in this thesis is partially supported by SMARTER Spanish project [33]. This research project aims to develop and evaluate novel imaging tools that can be integrated into the screening workflow to steer image acquisition and guide the selection of appropriate personalised screening protocols; and to process imaging data in an intelligent way to minimise interpretation time. The research herein aims to contribute to the overall aim of the SMARTER project by developing efficient methodologies to facilitate the use of CAD systems for breast cancer screening. The objectives of this research are summarized as:

- Propose a semi-automatic framework to perform an assessment of ABUS lesion volumes, as well as quantify the volumetric changes during lesions diagnosis and follow-up.

- Develop a patch-based mass detection framework on FFDM using CNNs, where small regions of images, i.e. patches are used for training and testing the CNNs.

- Propose a mass detection framework based on Faster R-CNN which takes whole FFDM as the input and provides the localisation of the lesions within this mammogram as an output.

1.7 Thesis Outline

In **chapter 1**, an overview of the breast cancer is provided along with different imaging modalities used in breast cancer screening, followed by the aims and objectives of this thesis. In the following **chapter 2**, a description of the traditional CAD systems is provided. In addition, a description of the datasets used in this thesis is presented, along with a comprehensive overview of the state-of-the-art methods in 3D ultrasound and mammography.

Segmentation is widely used in breast imaging to discriminate between normal and abnormal tissues (lesion) based on certain image properties. In **chapter 3**, a semi-automatic lesion segmentation framework for 3D ABUS volumes is presented. The localisation of lesion is done by radiologist by selecting a seed point (pixel) inside the 3D volume, and thereafter the 3D lesion volume is segmented around the seed point.

Following on from the works in 3D ultrasound images, the focus is shifted on 2D mammography images (using public datasets). This is done to facilitate the development of automated CAD systems based on emerging deep learning technology. In **chapter 4** a particular class of deep learning i.e. CNN is used for developing an automated CAD system for lesion detection using two available public mammography datasets. As the first step, a patch-based methodology is employed which uses small regions of the images (patch) to train the CNN. Later, in **chapter 5**, a lesion detection framework is presented which uses the whole mammogram for the training and testing of the CNN. This proposed framework overcomes one of the main limitations of the patch based approach i.e. computational efficiency. The **chapter 6** concludes research done in this thesis, highlights the contributions and outlines the areas of future research.

Chapter 2

Computer Aided Detection: Datasets and State of the Art

In the previous chapter, an overview of breast cancer, anatomy and different imaging modalities was provided. This chapter provides a description of the traditional CAD systems, the datasets used in this thesis and the state-of-the-art methods in 3D ultrasound and mammography.

2.1 Computer Aided Detection

In the last two decades, several Computer Aided Detection (CAD) systems have been developed to support radiologists for an early detection of masses in mammograms. However, these systems have limited effectiveness in terms of accuracy and are prone to reduce the number of False Positives (FP) and False Negatives (FN) [34, 35]. CAD systems have been an integral part of screening mammograms [36], despite questions about its effectiveness in the current form.

The CAD Systems can be divided into two main categories: Computer Aided Detection (CADe) and Computer Aided Diagnosis (CADx) mammographic systems. CADe systems indicate the presence of possible abnormalities whereas CADx systems classify potential lesions into malignant or benign in terms of malignancy likelihood. The CADe systems are used to search for abnormalities in the breast such as masses, micro-calcifications, architectural distortions or

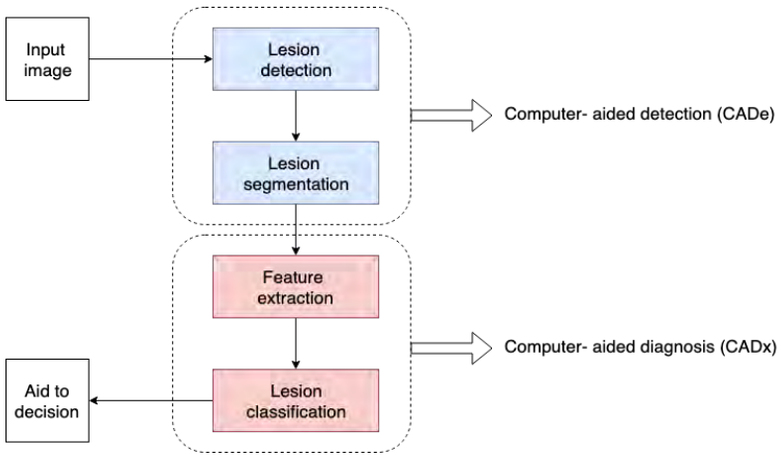


Figure 2.1: Flow chart showing different steps involved in CAD system and their characterisation into CADe and CADx.

asymmetries. A typical flow chart of the CAD system is shown in Fig. 2.1.

In previous research it has been shown that the early CADe systems had high sensitivity for macro-calcifications (up to 99%) [37], comparatively lower for masses (75-89%) [38] and least for architectural distortions (38%) [39]. The CADx systems are used to help radiologists to discriminate between benign and malignant lesions, and assist in the interpretation of positive mammograms (biopsy proven cancers). This task is usually a two-class classification problem. Current research is focused on the development of semi-automatic or fully automatic CADe/CADx systems for the detection and diagnosis of the breast lesions.

The images acquired using Ultrasound (US) Imaging technique contains speckle noise resulting in artefacts, thus limiting the accuracy of the CAD system. Automated Breast Ultrasound (ABUS) is the latest technology in US imaging modality, which produces the 3D scan of the breast. Since, in ABUS, the radiologist have to read large number of 2D slices, the main focus of CAD tools developed for ABUS, has been the effective reading of large ABUS volumes, and reducing the inter and intra-reader variability on manual evaluation [40–44]. In this thesis, efforts are made to develop a semi-automatic CAD tool to facilitate the reading of ABUS volumes by the radiologist.

Traditionally, CAD systems for mammography used Screen-Film Mammogram (SFM) with poor image quality, resulting in less robust application process. Moreover, they also relied on hand-engineered features like Hog, SIFT [45], which are not always consistent with respect to the Ground-Truth (GT), and have a human bias factor. In this thesis, efforts are also made to develop an automated CAD systems for detecting lesions in mammograms.

Table 2.1: Summary of the ABUS dataset.

	ABUS Dataset
No of volumes	56 (28 temporal pairs)
Malignant cases	26 (13 temporal pairs)
Benign cases	30 (15 temporal pairs)
Bit depth	8
Resolution	$0.11 \times 0.22 \times 0.53\ mm^3$

2.2 Datasets

In this section, description of the 3D Ultrasound and mammography datasets used in this thesis are presented.

2.2.1 3D Ultrasound

The 3D Automated Breast Ultrasound (ABUS) dataset used in this thesis is collected from the high-risk ABUS screening trials at Radboud University Medical Centre, Nijmegen, Netherlands between 2011 and 2014 using a Siemens ACUSON S2000 ABVS (Siemens Medical Solutions, Mountain View, CA, USA). This retrospective study was approved by the local institutional review board and the requirement for informed consent is waived.

The ABUS volumes were acquired from 15 patients (average age 48 ± 15 years) in three to five different views: anterior-posterior, medial and lateral (left and right), resulting in a total of 28 volumes. For these patients, depending on the nature of lesion i.e. benign (BI-RADS 2/3) and malignant (BI-RADS 4/5), another set of 28 ABUS volumes were acquired within a minimum time interval of 17 days to a maximum 2 years and 6 months. For a particular patient, based on the time of acquisition, different sets of temporal volumes were considered (one for each view) and referred in the text as prior and current. The obtained dataset contained all high-resolution anisotropic ABUS volumes and the details are summarised in Table 2.1. In Fig. 2.2 a sample of ABUS volume from the dataset is shown, with the GT rendered in 3D.

2.2.2 Mammography

Some of the publicly available mammography datasets include DDSM [47], CBIS-DDSM [48], INbreast [49], mini-MIAS [50], BCDR [51] etc. In this thesis, the two most popular and commonly used datasets: CBIS-DDSM and INbreast,

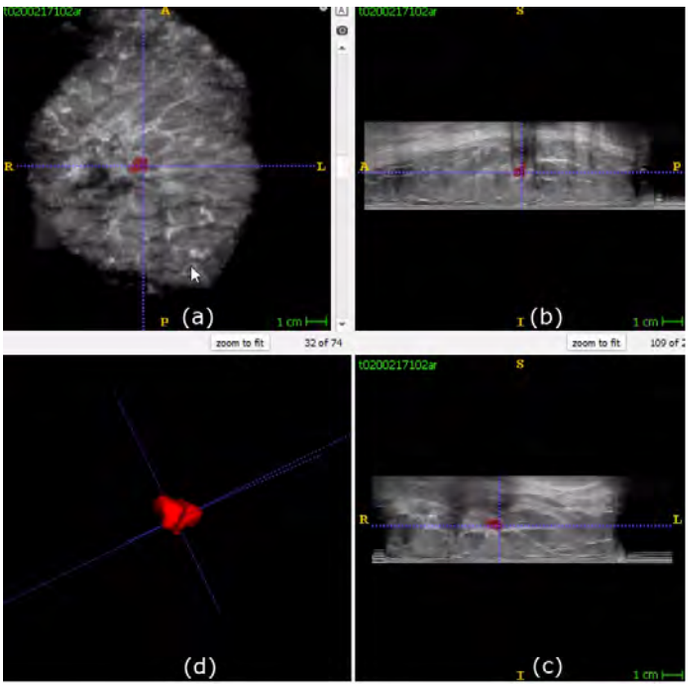


Figure 2.2: Sample of ABUS volume with ground-truth rendered using ITK-Snap [46], (a) coronal view, (b) sagittal view, (c) axial view, and (d) GT in 3D.

along with a large scale private dataset OMI-DB are used. Note that all the mammograms are available as "for presentation".

2.2.2.1 CBIS-DDSM

The DDSM [47] dataset contains approximately 2,500 studies. It contains SFM compressed with lossless JPEG encoding. In this thesis, a newer version of the dataset is used, i.e. CBIS-DDSM [48], containing a subset of the original DDSM images in the standard DICOM format. The dataset was downloaded on October 10, 2017 from CBIS-DDSM website [52] containing 3,061 mammograms of 1,597 cases. In total there are 1,698 masses in 1,592 images from 891 cases which includes both Cranio-Caudal (CC) and Medio-Lateral Oblique (MLO) views for most of the screened breasts. Figure 2.3 shows two mammographic views of the same case (CC and MLO) from the dataset.

2.2.2.2 INbreast

The INbreast [49] dataset is composed of Full-Field Digital Mammogram (FFDM) acquired using a Siemens MammoNovation mammography system (Siemens Healthineers, Erlangen, Germany). The images were acquired from 115 cases with CC and MLO breast views, leading to a total of 410 images available in DICOM

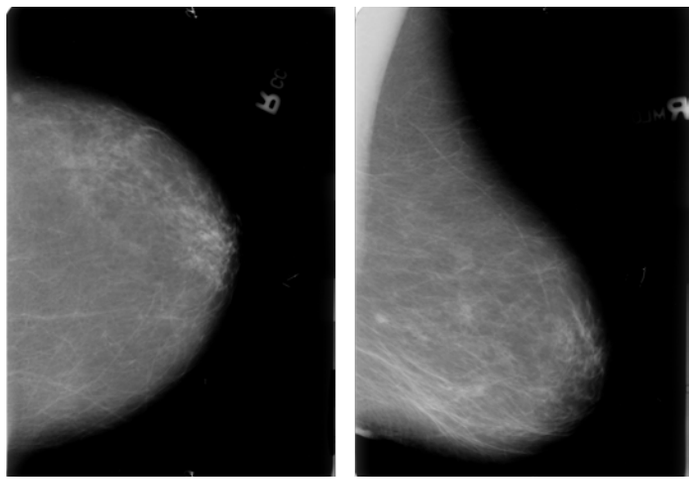


Figure 2.3: Sample SFM from CBIS-DDSM [48] showing two different views of the same case, left: CC and right: MLO.

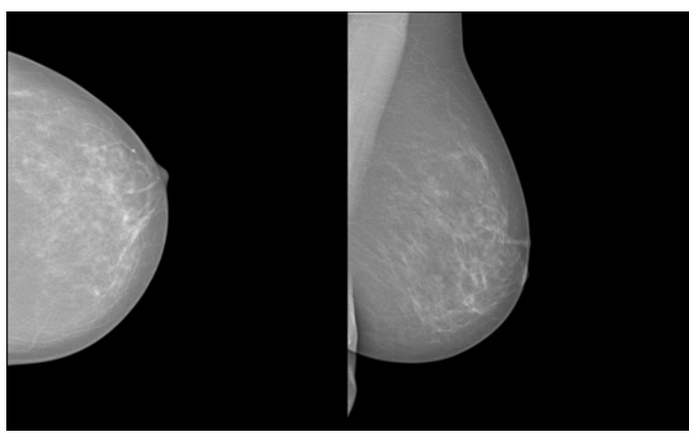


Figure 2.4: Sample FFDM from INbreast [49] showing two different views of the same case, left: CC and right: MLO.

format. From these, a total of 116 masses can be found in 107 mammograms from 50 cases. Figure 2.4 shows two views of the FFDM (CC and MLO) of the same case in the dataset.

2.2.2.3 OPTIMAM Mammography Image Database (OMI-DB)

The OMI-DB [22] is an extensive mammography image database of over 145,000 cases (over 2.4 million images) comprised of unprocessed and processed FFDM from the NHS Breast Screening Programme of the United Kingdom, which also contains expert's determined GT and associated clinical data linked to the images. We obtained a subset of this database comprising of 4750 cases. In this dataset there are images from fours different manufacturer including Hologic, Philips, General Electric (GE) and Siemens, containing a total of 2,419 cases with 4,217 masses and

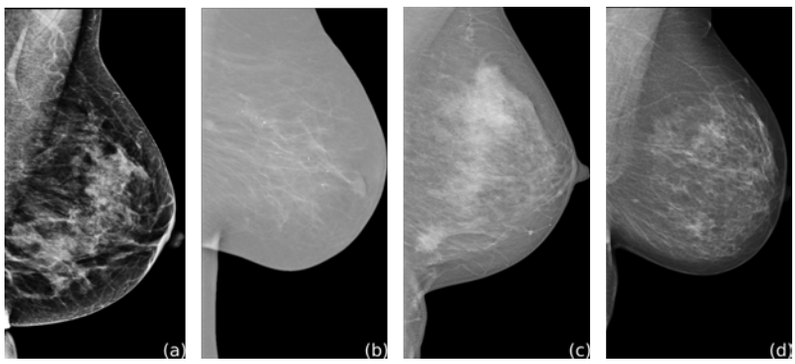


Figure 2.5: Sample FFDM from OMI-DB dataset from different manufacturers. Here mammograms are from (a) Hologic, (b) GE, (c) Siemens and (d) Philips.

Table 2.2: Summary of the mammography datasets including the details of scanner used for image acquisition, image sizes, resolution, bit depth and the total number of cases.

Dataset	Type	Scanner	Size	Resolution μm	Cases	Bit depth
CBIS-DDSM	SFM	Multi-Scanner	3000×4800	42, 43.5, 50	1597	8
INbreast	FFDM	Siemens	$\frac{3328 \times 4084}{2560 \times 3326}$	70	115	14
OMI-DB	FFDM	Hologic	$\frac{3328 \times 2560}{4096 \times 3328}$	60, 70	2884	12
		GE	$\frac{2294 \times 1914}{3062 \times 2394}$	100	207	12
		Philips	5355×4915	50, 83	242	12
		Siemens	4084×3328	70	32	12

946 cases without any mass. Figure 2.5 shows sample FFDMs from the scanners of different manufacturers in the OMI-DB dataset.
The summary of mammography datasets used in this thesis are shown in Table 2.2.

2.3 State of the Art

2.3.1 Automated Breast Ultrasound

Although many CAD algorithms have been explored in 2D Hand-Held Ultrasound (HHUS) [53–59], only a few studies have been proposed for 3D ABUS volumes. For instance, Chen et al. [40] used an active contour model to segment breast tumours in 3D images reconstructed from 2D ultrasound, while Moon et al. [42] used speckle and morphological features to classify breast masses in ABUS volumes showing a classification accuracy of 84.4%. One of the previous studies

on ABUS for breast cancer detection showed an increase in the diagnostic yield from 3.6 per 1000 with mammography alone to 7.2 per 1000 women screened with automated ultrasound [60].

Tao et al. [61] developed a breast lesion segmentation approach directly for ABUS using dynamic programming in combination with a spiral scanning technique (Dice Similarity Coefficient (DSC)$= 0.70 \pm 0.16$) and was extended for malignant lesion using a depth-guided dynamic programming method (DSC$= 0.73 \pm 0.14$) [62]. Later, Diaz et al. [63] extended the segmented method developed by Pons et al. [56] for 2D HHUS based on Markov Random Field-Maximum a Posteriori to 3D volumes, showing a decrease in performance from 0.75 to 0.55, in terms of DSC.

In recent years, there have been a lot of development in ABUS technology, both clinically and also in the field of CAD systems [64]. The main focus of CAD tools developed for ABUS has been to improve reading efficiency of large breast volumes reducing the inter and intra-reader variability. During diagnosis, when scans over time are available (temporal studies), it is also possible to track the lesion changes and see the effect of biopsy (if performed). However, this comes at the cost of longer reading and evaluation times, given the larger amount of images to be interpreted.

In breast imaging, segmentation is widely used to discriminate between normal and abnormal tissues based on certain image properties (e.g. intensity variation, texture). However, due to inherent speckle noise and low contrast of breast ultrasound images, automatic lesion segmentation is still a challenging task. Traditionally, the common segmentation algorithms described in the literature are based on: edge detection [65], Otsu threshold [66], region-based segmentation [67], morphological watersheds [67] etc. Each of these methods are based on different imaging principles, which might be effective for one particular application but may not be accurate for other.

In 2014, Lo et al. [68] used a specific type of Watershed (WAT) called Toboggan Watershed for whole ABUS image segmentation. In the author's previous work completed in 2016 [69], a semi-automatic framework was presented for the applicability of different segmentation algorithms from state-of-the-art on the lesion segmentation of ABUS volumes. The framework was semi-automatic in a way that the expert (radiologist) selects a point in the suspected region. Results suggested the use of WAT algorithm for ABUS lesion segmentation obtaining the best results with an average DSC of 0.70 ± 0.13 on a dataset of 56 ABUS volumes (30 benign and 26 malignant lesions).

2.3.2 Mammography

Following the recent developments in computer technology and data sciences, there has been a lot of interest in exploring deep learning methods for mammography [31, 70–74]. The term "Deep Learning" can be defined as any one of a set of methods that learn data representations using multiple levels of representation [75]. They are obtained by composing simple but non-linear models that transform the representation from one level (starting with the raw input) into increasing levels of representation. Deep learning strategies have recently gained a lot of interest in various fields including object detection [75–79], image recognition [77–83], natural language processing [84, 85] and speech recognition [86, 87]. In Deep Learning, Convolutional Neural Network (CNN) is most commonly used to analyse images.

Several authors have proposed the use of traditional machine learning and content based image feature retrieval techniques to classify masses and micro-calcifications [88, 89]. The feature exploitation of deep learning frameworks in the field of breast imaging has been limited, as only a small number of public datasets are available (e.g. DDSM [47], INbreast [49]). In the early paper of Kozegar et al. [90] in 2013, the authors used an iterative breast segmentation approach to subsequently classify the regions using traditional feature selection and machine learning paradigms.

In 2015, Dhungel et al. [70] proposed a multi-scale deep belief network (m-DBN) classifier, followed by a cascade of Region-Based CNN (R-CNN) and cascades of random forest classifiers for mass detection, obtaining True Positive Rate (TPR) of 0.96 ± 0.03 at 1.2 False Positives per Image (FPI) on INbreast and 0.75 at 4.8 FPI on DDSM-BCRP dataset [47]. Later, Carneiro et al. [71] proposed the use of CNN models pre-trained using a computer vision database (ImageNet) for classifying benign and malignant lesions obtaining an Area Under Receiver Operating Curve (AUROC) of 0.91 ± 0.05 on INbreast and 0.97 ± 0.03 on DDSM dataset.

More recently in 2017, Lotter et al. [72] trained a patch-based CNN to classify lesions in the DDSM dataset and subsequently used a scanning-window approach to provide full mammogram classification achieving an AUROC of 0.92 on the DDSM dataset. In scanning-window approach, the image is partitioned into set of patches (with minimal overlap), such that each patch is contained completely within the image and the entire image is completely tiled. Later, Dhungel et al. [31] used a deep learning methodology to develop an approach for mass detection, segmentation and classification in mammograms and tested the approach on the

INbreast dataset. Detection results had a TPR of 0.90 ± 0.02 at 1.3 FPI on testing data.

Regarding the use of private mammography datasets, Becker et al. [91] used a multi-purpose image analysis software and an internal database from 3,228 patients to train the Artificial Neural Network (ANN). The model was then tested on the BCDR dataset [51] of 35 patients to obtain AUROC of 0.79. In other works, Kooi et al. [73] used a larger private database of $\approx$45,000 FFDM to provide a comparison between traditional mammography CAD systems relying on hand-crafted features and the CNN methods. It was shown that the CNN model trained on a patch level with a large database outperformed state-of-the-art CAD systems, obtaining an AUROC of 0.929 compared to 0.906 obtained for the CAD system.

Researchers have used Faster R-CNN in medical imaging [92, 93], but a very small amount of literature is available in the field of breast imaging. For instance, Akselrod-Ballin et al. [94] used a modified version of Faster R-CNN model to include informations from the finer bottom levels during classification stage. Results were then evaluated on an internal database of 850 images to obtain AUROC of 0.78. Later, in [95], the results were evaluated on a subset of the INbreast dataset (with masses) obtaining a TPR of 0.93 at 0.56 FPI. Ribli et al. [74] trained a Faster R-CNN on the DDSM database composed of 2,620 SFM, and then evaluated the performance of the network on the INbreast database of malignant lesions, obtaining a TPR of 0.90 at 0.3 FPI and AUROC of 0.95.

Jung et al. [98] proposed a mass detection model based on RetinaNet [101] and used a new loss function called focal loss to address the problem of extreme class imbalance between foreground and background. The performance of the network was evaluated on the INbreast to obtain the best TPR of 0.97 ± 0.05 at 3.0 FPI and 0.94 ± 0.02 at 1.3 FPI. Morrel et al. [99] presented a neural network based on region-based fully convolutional network (R-FCN) and deformable convolutional nets. The network was trained using the OMI-DB [22] dataset, achieving AUROC of 0.867 for breast-wise detection in the DREAM challenge on 13,000 images from Group Health.

Al-masni et al. [100] adopted the You Only Look Once (YOLO) deep learning method [102] for the detection and classification of masses in mammograms. The results showed an overall mass detection accuracy of 96.33% and the classification accuracy of 85.52% on the DDSM dataset. Although, authors showed that enhanced accuracies were obtained when using an augmented DDSM database created by rotating each mammogram multiple times. Authors used images of the same patient for both training and testing instead of splitting training and testing

Table 2.3: Summary of other works on mass detection in mammography using deep learning methods.

Author (Year)	Dataset (Images)	Method	Mammogram size	AUROC	TPR at FPI
Kozegar et al. [90] (2013)	mini-MIAS (330)	Iterated segmention	512×512	n/a	0.91 at 4.8
	INbreast (116)				0.87 at 3.67
Carneiro et al. [71] (2015)	INbreast (410)	Fast CNN (CNN-F) [96]	264 × 264	0.91±0.05	n/a
	DDSM (680)			0.97±0.03	
Dhungel et al. [70] (2015)	INbreast (410)	m-DBN+R-CNN+RF	264×264	n/a	0.96±0.03 at 1.2
	DDSM-BCRP (316)				0.75 at 4.8
Lotter et al. [72] (2017)	DDSM (10480)	InceptionV3	patch (256×256)	0.77 ± 0.03	n/a
		wide ResNet [97]		0.92 ± 0.02	
Dhungel et al. [31] (2017)	INbreast (410)	cascade R-CNN+RF with hypothesis refinement	264×264	n/a	0.90 ± 0.02 at 1.3
Becker et al. [91] (2017)	BCDR (70)	ANN using ViDi software	high resolution	0.79	n/a
Kooi et al. [73] (2017)	internal(45000)	downscaled VGG	patch (250×250)	0.929	n/a
Akselrod-Ballin et al. [94] (2017)	internal (850)	modified Faster R-CNN	grid (800×800)	0.78	n/a
Akselrod-Ballin et al. [95] (2017)	internal (750)	modified Faster R-CNN	grid (800×800)	n/a	0.9 at 1.0
	INbreast (100)				0.93 at 0.56
Ribli et al. [74] (2018)	INbreast (n/a)	Faster R-CNN	2100 × 1700	0.95	0.90 at 0.3
Jung et al. [98] (2018)	GURO (222)	RetinaNet	n/a	n/a	0.98 ± 0.02 at 1.3
	INbreast (410)				0.94 ± 0.05 at 1.3
Morrel et al. [99] (2018)	Dream challenge (13,000)	R-FCN/DCN	2,545 × 2,545	0.8667	n/a
Al-masni et al. [100] (2018)	DDSM (600)	YOLO	448×448	0.877	n/a
Al-antari et al. [32] (2018)	INbreast (410)	YOLO	448×448	0.948	n/a

datasets at a patient level. This type of the distribution of training and testing images would result in mammograms from the same patient to be present in both training and testing, thus giving potentially biased results. In other works, Al-antari et al. [32] presented a fully integrated CAD system adding lesion segmentation to the framework proposed in [100]. The results were presented on the INbreast dataset, but again the augmented dataset is used and the distribution is made based only on the images raising issue on overlap between training and testing patients.

A summary of the works performing mass detection in mammography are summarized in Table 2.3.

Chapter 3

Lesion Segmentation in Automated Breast Ultrasound

In the previous chapter, an overview of the Computer Aided Detection (CAD) systems was provided along with the datasets and state-of-the-art methods. This chapter deals with one of the steps in the CAD systems i.e. lesion segmentation (see Fig. 2.1). Segmentation is widely used in breast imaging to discriminate between normal and abnormal tissues (lesion) based on certain image properties (e.g. intensity variation, texture, etc.). Herein, a semi-automatic lesion segmentation framework for 3D Automated Breast Ultrasound (ABUS) volumes is presented.

3.1 Introduction

Ultrasound imaging is the standard technology used for cancer screening in dense breast. The conventional Ultrasound (US) or Hand-Held Ultrasound (HHUS) have a major disadvantage: poor quality because of multiplicative speckle noise that results in artefacts and requires highly skilled operators to detect lesions [103]. Moreover, due to the disturbances during scans (patients movement, breathing effects), sometimes multiple scans are required to be done which makes it a time-consuming process and is also prone to human error.

ABUS is an alternative procedure to perform ultrasound of the breast using an automated procedure to overcome the time-consuming process and operator

dependence nature of hand-held procedure [104]. Using ABUS, an exhaustive scan of the whole breast is produced using 3D technology, and only requires the basic knowledge of the set-up to obtain a good scan of the breast. One of the limitations of ABUS is that the radiologists have to analyse each slice of the 3D volume to identify suspicious lesions.

Lesion change is typically analysed by measuring the longest lesion axis in three dimensions and the lesion volume is estimated using this information, without accounting for more accurate volumetric calculation [105]. The lesion volume analysis has been used in other areas of medical imaging such as Multiple Sclerosis lesion in brain [106–108] and lung CT images [109]. It has been also shown in the literature that the lesion volume measurements can be used to quantify the disease progression [109, 110].

Image segmentation plays an important role in Medical Image Analysis as it is one of the crucial steps in many clinical applications such as detection of tumours, veins, organs etc. Selection of a segmentation algorithm, truly depends on the type of images, application and requirements. In addition, accurate segmentation of lesions is a challenging problem for many reasons such as: lesions have considerable variation in shape, have an overlapping area with normal tissues which is difficult to differentiate and distribution of intensities in the lesions are very high.

In this chapter, investigations are done to analyse the impact of different de-noising algorithms on segmentation results. Moreover, a temporal volumetric assessment of breast lesion changes using a temporal dataset is presented. From the clinical perspective, this could be used to provide valuable information to radiologists about the changes in lesions before and after a surgery or therapy is performed.

3.2 Methodology

The lesion segmentation framework proposed here is described in Fig. 3.1. The input to the framework is the 3D ABUS volume and a seed point (in Cartesian coordinates) which corresponds to a voxel within the lesion region. Herein, the seed points are located manually by selecting one of the most inner pixels in the central slice of the lesion following expert's annotations. The output of the framework is the binary 3D lesion segmentation and its estimated volume. In the following sub-sections, details of the steps of the proposed framework are presented.

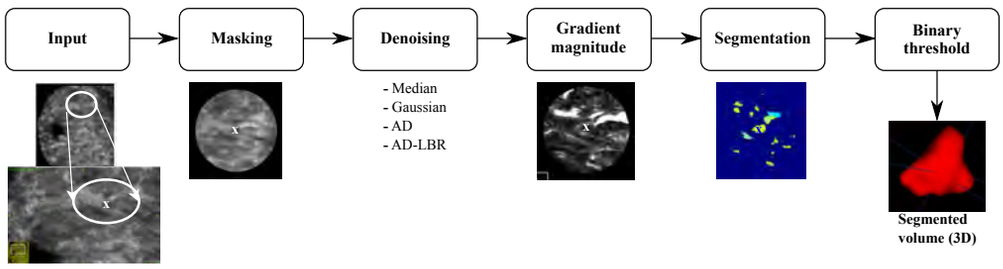


Figure 3.1: The proposed segmentation framework, the corresponding seed point is shown with a 'x' marker. Analysed de-noising methods are listed under the de-noising block.

3.2.1 Masking

The acquired ABUS images are high-resolution 3D volumes which makes the segmentation process computationally expensive. Hence, all ABUS volumes are down-sampled from their original resolution $0.11\times0.22\times0.53$ mm^3 (2.1) to $0.6\times0.6\times0.6$ mm^3 isotropic voxels (uniform size in all the three directions). Since the new voxel resolution is not a multiple of the original resolution, linear interpolation is used to estimate the new voxel's intensity values.

As the volume of the lesion is small compared to the entire ABUS volume, a masking step is added to speed-up the processing algorithms. This is performed to include enough distance from the lesion's border to the edge of the mask to avoid edge artefacts in the volume of interest. Assuming an oval or elliptical shape of lesion [111, 112], the central slice of the prior and current lesion showed an average major axis of 10.68 ± 5.23 mm and 10.0 ± 4.53 mm, respectively. Thus to include all the lesions within the chosen dataset, the volumes are masked using a sphere of 75 mm radius centred at the seed point. A sample case of the masking process is illustrated in Fig. 3.2.

3.2.2 De-noising

Similar to conventional US, ABUS volumes suffer from low contrast and speckle noise patterns which limits the efficacy of posterior analysis steps. Therefore, it would be advantageous to perform suitable de-noising of ABUS volumes before applying any segmentation algorithm. In this thesis, investigations are made to analyse the impact of different de-noising algorithms including median filtering, Gaussian filtering, Anisotropic Diffusion (AD) and Anisotropic Diffusion-Lattice based Reduction (AD-LBR). Herein, all the de-noising algorithms in 3D are analysed using implementations in Insight ToolKit (ITK) [113]. A brief description of these methods are as follows:

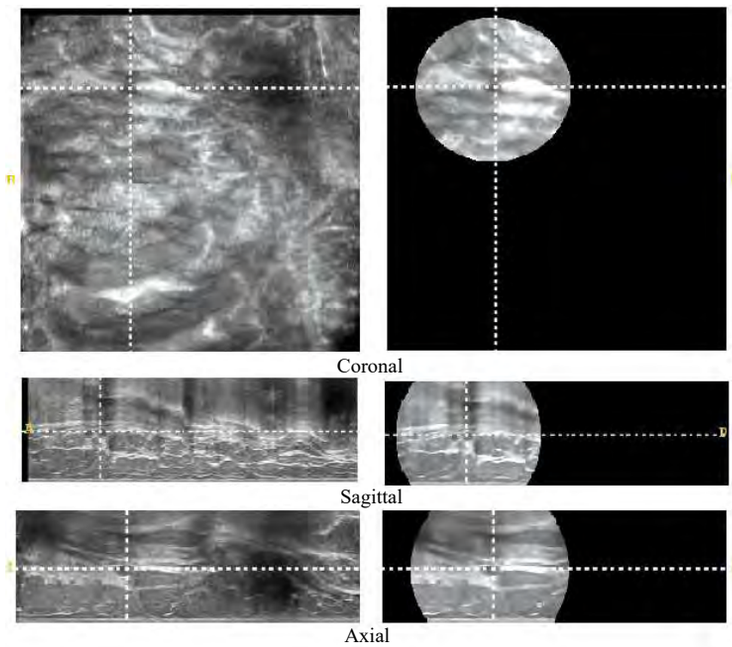


Figure 3.2: ABUS volume (0.6 mm^3 pixel spacing) with dimensions: $254 \times 278 \times 74$ pixels and seed point inside the lesion as shown with the cross-hair on the left and the masked volume with the centre of the sphere at the seed point, with a diameter of 75 mm on the right. Three different views are shown as Coronal (top), Sagittal (middle) and Axial (bottom).

Median filtering: It is a non-linear digital filter particularly used for salt and pepper noise. The ITK implementation with a 3D filter of radius $3 \times 3 \times 3$ pixels is used in this chapter [114].

Gaussian filtering: In this filter it is assumed that the noise present in the image follows a Gaussian distribution and hence a Gaussian Kernel is used to perform de-noising in three directions [115, 116]. In this work a Gaussian Kernel of standard deviation $\sigma_x = \sigma_y = \sigma_z = 0.6$ mm is used.

AD: The motivation for AD also called non-uniform or variable conductance diffusion is that a Gaussian smoothed image is a single-time slice of the solution to the heat equation, and the idea is to diffuse the pixel intensities preserving the sharp boundaries (details) of the image. In here, the 3D extension of the classic Perona-Malik anisotropic diffusion equation for scalar-valued images [117] is used, with the conductance ($c = 3$) and iterations ($i = 20$) controlling the extent of de-noising.

AD-LBR: Weickert [118] showed that a filter can be locally adapted such that it is truly anisotropic close to linear structures like edges or lines and also is elongated along the structure and narrow across. These methods are classified

as Edge Enhancing Diffusion (EED) or Coherence Enhancing Diffusion (CED). Later, Fehrenbach et al. [119] introduced AD-LBR to discretize AD based on Lattice Basis Reduction (LBR). CED heuristically smooths across all image except contours, while EED smooths only at image locations which are tangential to the image contours. A detailed description of the mathematical formulation is given by Mirebeau et al. [120]. In the non-linear diffusion algorithm, a structure tensor S is used as an estimator of the gradient direction of an image u, which depends on two parameters i.e. the noise scale σ, and the feature scale ρ.

$$S_u := K_\rho * (\nabla u_\sigma \nabla u_\sigma^T) \qquad where \qquad u_\sigma := K_\sigma * u \tag{3.1}$$

where K is the Gaussian convolution. Weickert's diffusion tensor $D = D_u(x,t)$, is defined in terms of the eigen analysis of the structure tensor $S = S_u(x,t)$ as:

$$D = \sum_{1 \le i \le d} \mu_i e_i \otimes e_i \qquad where \qquad S = \sum_{1 \le i \le d} \lambda_i e_i \otimes e_i \tag{3.2}$$

where $d \in \{2,3\}$ is the image dimension, λ_i are the eigenvalues of the structure tensor, with $0 \le i < d$. The diffusion tensor D is defined by the same eigenvectors, but with modified eigenvalues μ_i. EED and CED can now be defined by Eq. 3.3 and Eq. 3.4.

$$\mu_i = 1 - (1-\alpha) exp\Big(- \Big(\frac{\lambda}{\lambda_i - \lambda_1} \Big)^m \Big), \tag{3.3}$$

$$\mu_i = \alpha + (1-\alpha) exp\Big(- \Big(\frac{\lambda}{\lambda_d - \lambda_i} \Big)^m \Big) \tag{3.4}$$

In this work the feature scale ρ is set at 2, while the range of other parameters were empirically selected to maximise signal to noise ratio [121] for each image as: diffusion time$\in [1,2]$, lambda$(\lambda) \in [0.001, 0.1]$ noise-scale$(\sigma) \in [0.1, 1]$.

The results for different de-noising methods are shown in Fig. 3.3.

3.2.3 Segmentation

Following on from the author's previous work [69], in this thesis Watershed (WAT) algorithm is used for ABUS lesion segmentation. WAT is a morphological method

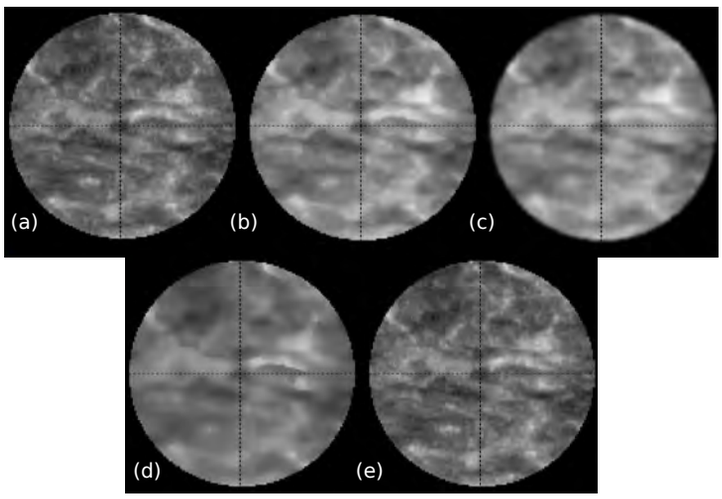


Figure 3.3: A qualitative example of the different image de-noising methods. a) original image without de-noising, b) Median filter (r=3x3x3 pixels), c) Gaussian smoothing ($\sigma_x = \sigma_y = \sigma_z = 0.6$mm), d) AD ($C = 3, i = 20, \delta t = 0.0624$), e) AD-LBR ($t = 1.5, \lambda = 0.005, \sigma = 0.8, \rho = 2$).

used extensively in the field of image processing. The underlying theory of the approach is well documented in literature [121–123] and a brief overview is provided here. The implementation of WAT includes creation of a topological map based on the pixel's grey level value, e.g. a white pixel represents peak, black pixel represents valley and all other intermediate pixels can be distributed accordingly. As a result, several 'catchment basins' are formed corresponding to different local minima. The two adjacent basins are separated by rigid lines to form two separate regions in the image called watershed regions as shown in Fig. 3.4.
In this chapter, ITK implementation of WAT segmentation is used with the following parameters:

Level: It controls the watershed depth (0-1 of the maximum depth of the input image). The level parameter is dynamically controlled, by moving from the highest level to the lowest level in small increment of 0.1.

Threshold: It is specified as a fraction (0-1) of the maximum height of the image. Thresholding the minimum values in the image decreases the number of local minima and produces an initial segmentation with fewer segments. Empirical experiments found the threshold parameter of the filter does not play a very significant role in the output, and is fixed to a typical value of 0.005, as stated in the literature [63].

The final 3D segmentation volume corresponds to the binary region of the WAT algorithm which contains the seed point.

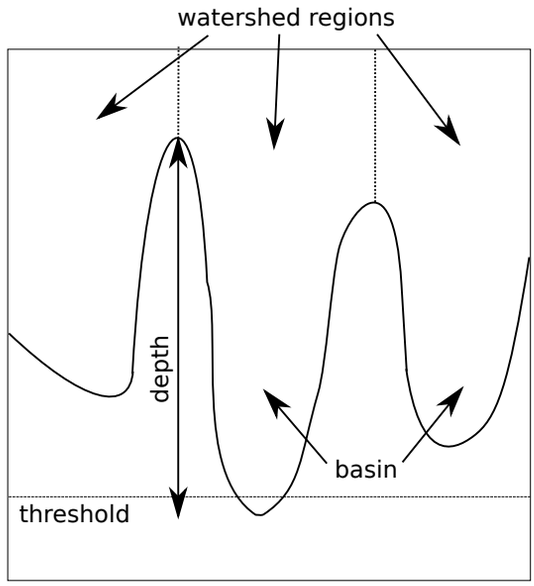


Figure 3.4: A schematic overview of watershed in 1D.

3.2.4 Evaluation Metrics

In order to evaluate the performance of the proposed framework, an experienced radiologist with more than 4 years of experience in ABUS, performed localisation and manual segmentation on a coronal 2D slice containing the lesion. This initial segmentation is expanded to a 3D volume using the open source ITK-Snap [124] software by medical imaging experts. Furthermore, the 3D volumes are verified (and modified when needed) by another experienced radiologist (25 years in breast imaging) to produce the final Ground-Truth (GT) volume. All the lesions are histologically proven from the biopsy and the categorisation is used to evaluate the performance of WAT segmentation algorithm.

Dice Similarity Coefficient (DSC) is a commonly used measure to provide information regarding the extent of overlap between two areas or volumes [125, 126] and is typically measured on the scale of [0-1], where 1 represents complete overlap. Herein, DSC is measured as

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{3.5}$$

Figure 3.5 illustrates the definitions used in evaluation measures.

- False Positives (FP): Voxels belonging to the lesion in the segmentation but not part of lesion in GT volume.
- True Positives (TP): Voxels corresponding to lesion in both segmentation and GT volume.

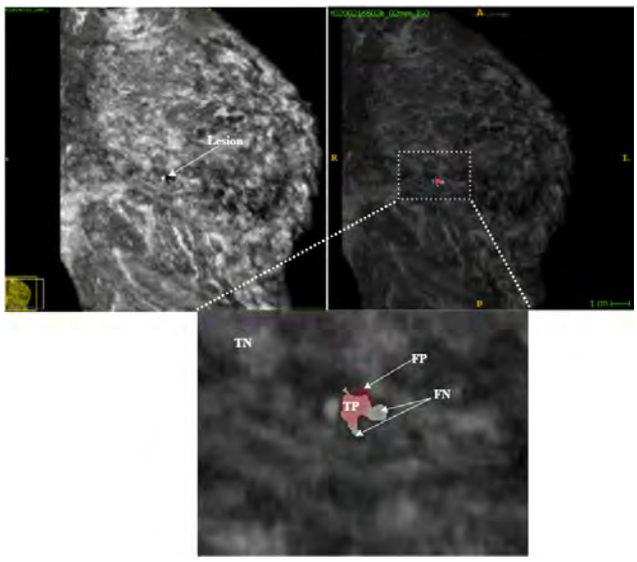


Figure 3.5: In the top left, the original image with lesion is shown (pointed lesion area). In the top right image, the volume is shown to overlay with GT (grey) and segmentation (red). In the lower image, segmentation results overlay with GT is shown, where TP is the intersection of grey and red region.

- False Negatives (FN): Voxels belonging to the lesion in GT but not included in the segmented volume.
- True Negatives (TN): All the voxels outside the lesion in GT and segmentation output.

The volumetric measure is another metric used to analyse the performance of the proposed segmentation framework. Since during the acquisition of ABUS volumes a small compression is used to stabilise the breast, it is assumed that the lesion volumes remain fairly constant during this process. This makes the lesion volume an important parameter from the radiologist point of view, especially for analysing the temporal evolution of the lesion.

The lesion volumes (in mm^3) for both GT (V_{GT}) and segmentation (V_{Seg}) are computed and a regression analysis is performed using Pearson correlation coefficient to evaluate the extent of similarity between the two volumes. Temporal analysis of the change in volumes (ΔV^t) is evaluated using the relative volume difference between prior and current ABUS acquisitions of the same case (Eq. 3.6).

$$\Delta V_i^t = \left| \frac{V_i^c - V_i^p}{V_i^p} \right| \times 100 \quad [\%], \tag{3.6}$$

where t, p, c represent temporal, prior and current volumes and i represent either Seg or GT. The proposed framework is developed using the open source ITK

segmentation and registration toolkit [113] libraries, and all the image processing routines are implemented in C++. All the statistical analysis is performed using one-pair sample test (ttest) in Matlab 2016a (MathWorks, MA, USA).

3.3 Experimental Results

In chapter 2, an overview of the ABUS dataset has been provided. In this chapter, the lesion segmentation is performed on 56 ABUS volumes of 0.6 mm^3 isotropic voxel spacing.

3.3.1 Lesion Segmentation

As the first step, the performance of different de-noising algorithms is evaluated by comparing WAT segmentation outputs (in terms of DSC, FP, FN). The obtained results are shown in Table 3.1, where μ and σ refer to the mean and standard deviation respectively, for all the volumes in the dataset. It is observed that AD-LBR outperforms other de-noising algorithms based on the evaluation measures used. This can be explained by the fact that AD-LBR performs smoothing over the volume considering the structure tensor i.e. location and direction of the gradient value, resulting in optimum smoothing and also preserving the edge information. These edge informations complement the WAT algorithm in segmenting 3D volumes. WAT segmentation with AD-LBR also resulted in lower FPs and FNs compared to other de-noising algorithms. These differences are statistically significant ($p < 0.05$) compared to no de-noising, median, Gaussian and conventional AD de-noising.

The box-plot in Fig. 3.6 shows that the DSC values obtained for the majority of cases using AD-LBR are in the range 0.69 ± 0.11, whereas other de-noising algorithms show a wider range of values. Moreover, there are many outliers in the results for AD de-noising, while using AD-LBR there are almost no outliers. Hence, all further analysis is performed using AD-LBR (for de-noising) with WAT segmentation.

3.3.1.1 Seed Point Selection

Since the proposed tool is semi-automatic, it requires a manual selection of a voxel within the lesion (i.e. seed point) prior to the WAT segmentation. For this reason, an additional analysis was performed to check the sensitivity of the seed point selection with the proposed method (AD-LBR + WAT). Segmentation results in terms of

Table 3.1: Results for different de-noising methods with WAT segmentation where μ and σ refer to the mean and standard deviation respectively.

Filters	DSC ($\mu \pm \sigma$)	FP ($\mu \pm \sigma$)	FN ($\mu \pm \sigma$)
No filter	0.55 ± 0.22	0.41 ± 0.24	0.39 ± 0.25
Median	0.64 ± 0.16	0.39 ± 0.19	0.25 ± 0.20
Gaussian	0.61 ± 0.19	0.40 ± 0.19	0.28 ± 0.24
AD	0.64 ± 0.20	0.38 ± 0.20	0.28 ± 0.24
AD-LBR	$\mathbf{0.69 \pm 0.11}$	$\mathbf{0.35 \pm 0.14}$	$\mathbf{0.23 \pm 0.15}$

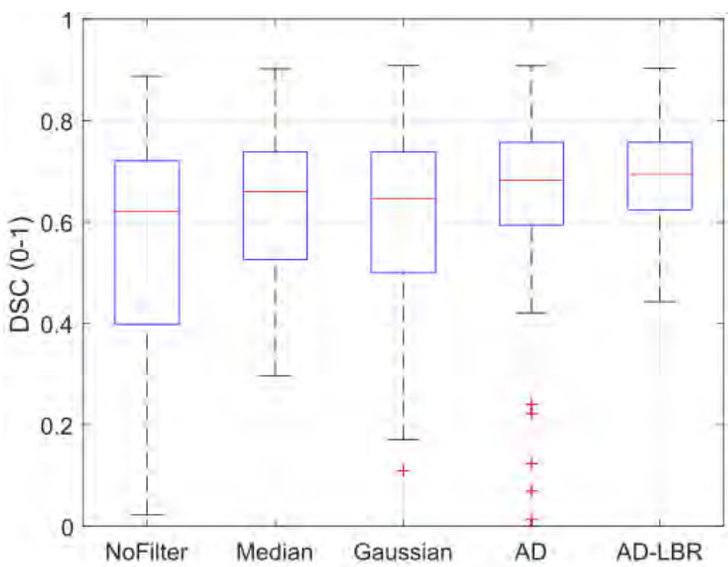


Figure 3.6: Box Plot summarising the performance of different de-noising algorithms on lesion segmentation.

DSC were evaluated for 15 different seed points randomly located within the lesion area, avoiding the boundaries for all the volumes. The resulting average DSC was 0.64 ± 0.10, showing a small variability of the proposed segmentation framework when selecting the seed point location randomly. So it could be concluded that the results obtained using the formulated framework are less sensitive to the selection of seed point. This helps to limit the operator dependency, as different radiologists may select different seed points.

3.3.2 Qualitative Analysis

As a qualitative result, lesion segmentation performed using the proposed framework for benign and malignant ABUS cases (coronal view) are shown in Fig. 3.7. Since all lesions are rated by an expert radiologist according to the BI-RADS classification, a separate analysis for benign (30 volumes) and malignant

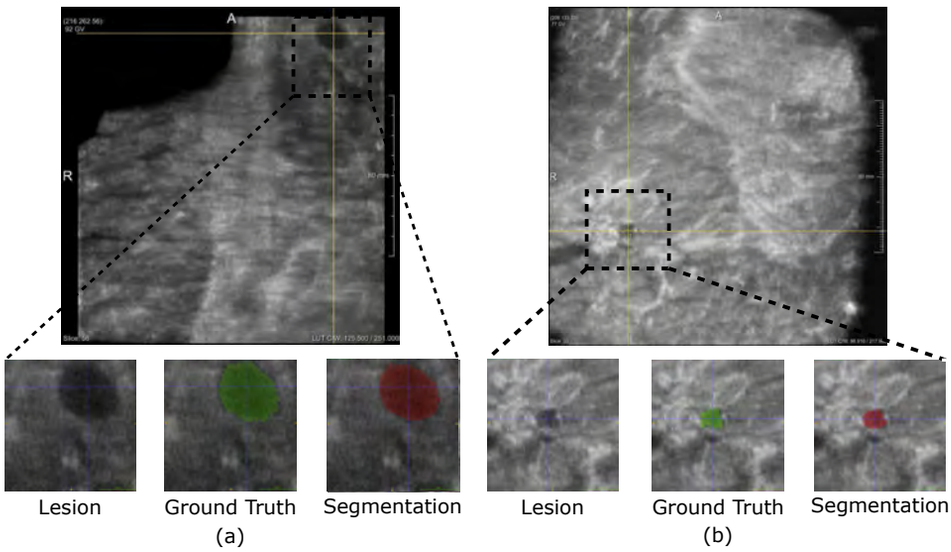


Figure 3.7: Lesion segmentation: (a) Benign (DSC = 0.90; FP = 0.04; FN=0.14), (b) Malign (DSC = 0.85; FP = 0.11; FN=0.18).

(26 volumes) cases is also performed to see if there is any segmentation limitation based on the malignancy of the lesions in the segmentation results. For benign cases the DSC$= 0.71 \pm 0.11$, while for malignant cases a slight decrease in DSC$= 0.66 \pm 0.11$ is observed. This could be due to the fact that the boundaries of the benign lesion are well defined and generally had uniform growth, while the boundaries of malignant lesions are irregular and mostly grow in speculated and non-uniform manner.

3.3.3 Volumetric Analysis

The volumetric analysis of the output of the proposed segmentation framework is performed, with the V_{GT} compared to the V_{seg}, including all the prior and current cases. The volumetric correlation plot is shown in Fig. 3.8, where a high correlation between V_{GT} and V_{seg} ($r^2 = 0.960$, $p < 0.05$) is obtained. This substantiates the applicability of the implemented framework for volumetric assessment in the used dataset. In order to quantify the agreement between the V_{GT} and V_{seg}, the Bland-Altman plot is shown in Fig. 3.9, where one can better perceive the correlation between the two results. Here the mean difference is 38 mm^3 with 95% confidence interval -151 to 230 mm^3, the variability of the volume differences can also be noticed.

A volumetric temporal analysis is then performed on the temporal ABUS volumes (28 prior and 28 current) to evaluate the changes in volume for both V_{GT} and V_{Seg} for prior and current cases respectively. The V_{GT} for prior cases

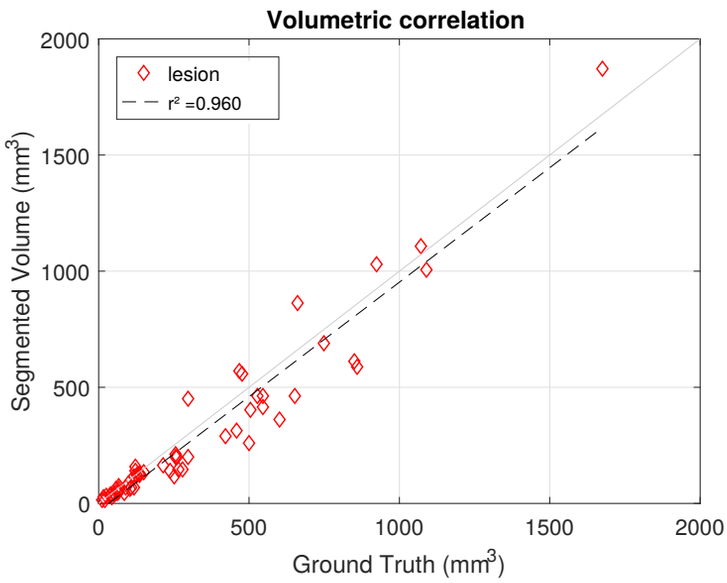


Figure 3.8: Volumetric correlation between Segmentation (V_{Seg}) results and Ground-Truth (V_{GT}).

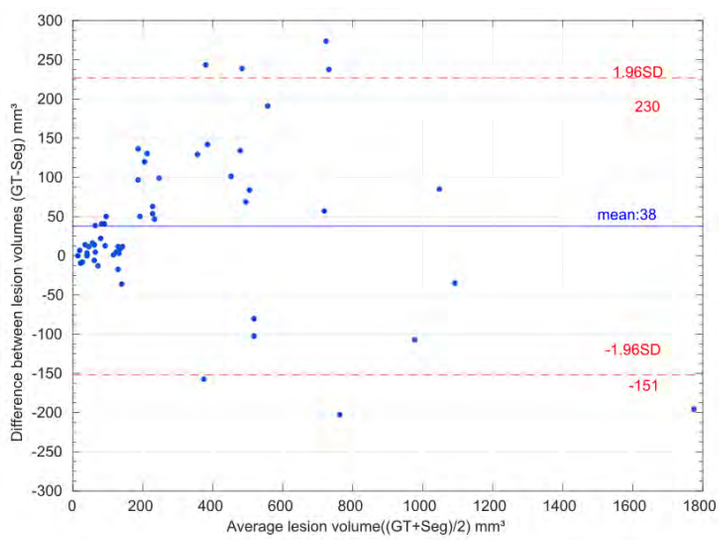


Figure 3.9: Bland Altman Plot of V_{GT} and V_{Seg} (in mm^3).

ranged from 18 mm^3 to 1148 mm^3, while for current cases V_{GT} ranged from 13 mm^3 to 1676 mm^3. In the same context, V_{seg} ranged from 16 mm^3 to 1109 mm^3 for prior cases and from 14 mm^3 to 1872 mm^3 for current cases. This showed that both GT and segmentation volumes follow a similar trend, which is complemented by the correlation result for both prior ($r^2 = 0.967, p < 0.05$) and current ($r^2 = 0.956, p < 0.05$) volumes as shown in Fig. 3.10.

It can be observed that for some lesions the framework showed a decrease in volume in the temporal study. These cases are individually analysed and it is found that these are malignant cases where a biopsy was performed between the two scans.

Therefore, the framework is able to capture the decrease in lesion volume from prior to current, but due to unavailability of the pathological biopsy information, the volumetric differences could not be verified.

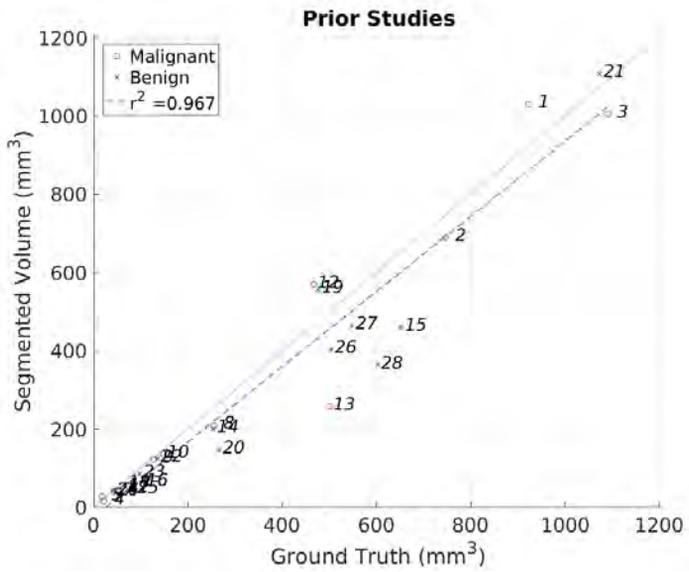
Finally, applicability of the proposed framework for the temporal analysis of breast lesion segmentation is substantiated by performing an error analysis. The obtained results showed that the temporal variation between the changes in GT and segmented volumes is not statistically significant ($p > 0.05$), indicating the potential of the framework to be used for temporal analysis.

3.4 Discussions and Conclusions

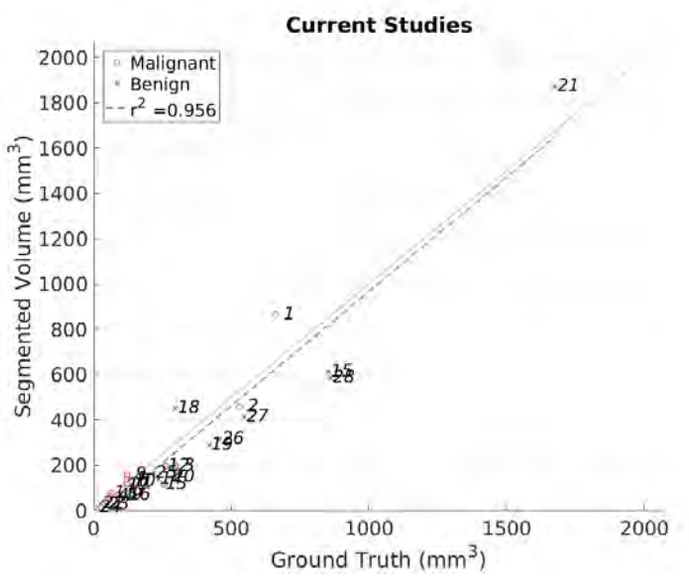
In this chapter, a breast lesion segmentation framework was proposed for ABUS volumes which provides support for radiologist's diagnosis in breast cancer examinations. The main contribution of this chapter is the development of a semi-automatic framework to segment lesion in ABUS, that could also be used for temporal analysis of breast volumes. This is clearly shown by the fact that the temporal variation in the segmented volumes (ΔV^t_{Seg}) is not significantly different ($p > 0.05$) to the variation in the GT volumes (ΔV^t_{GT}) and the segmented volumes have good correlation with the GT volumes ($r^2 = 0.960$).

Although lesion volume is used as a measure of the accuracy of the proposed segmentation algorithms, the potential applications of knowing such 3D information of the lesion are large: volumetric analysis for accurate knowledge of temporal lesion changes, feature extraction (e.g. texture) for lesion characterisation, etc. The effect of different de-noising algorithms on lesion segmentation for ABUS volumes was also explored. The results showed a significant difference in the segmentation output when WAT segmentation algorithm was used with different de-noising algorithms, in which AD-LBR outperformed ($p < 0.05$) all other methods for ABUS volumes with respect to evaluation measure (DSC, FP, FN) used.

In section 3.3.3, the lesion was segmented to estimate its volume, and subsequently compared with GT volume to measure the accuracy using Pearson correlation coefficient. The results of this volumetric analysis of the lesions showed a high correlation between GT and segmented volumes ($r^2 = 0.960, p < 0.05$). The average DSC obtained was 0.72 for benign lesion and 0.66 for malignant lesion, with a maximum of 0.90 (benign lesion) and minimum of 0.44 (malignant lesion), showing the challenging task of the segmentation algorithm. It was also seen from Fig. 3.8 that for very few cases the segmented lesion volumes was



(a)



(b)

Figure 3.10: Volumetric correlation between the V_{Seg} and V_{GT} for temporal study, (a) prior studies, (b) current studies. The numeric labels signify the same case in prior and current studies. The scale for the plots is different for the better visualisation.

greater than the GT volume, signifying that the proposed framework does not result in over segmentation of the lesions. These results demonstrate the potential of the framework to be used clinically for breast lesion segmentation and volume estimation.

In general, the temporal lesion volume change was similar for both the GT and the segmentation. In 2 out of 28 volume pairs, the trend between prior and

current cases for GT and segmentation was different and it was found that the segmentation showed a decrease in the volume, whereas the GT showed an increase in the volume. This issue requires further analysis but could be attributed to specific lesion properties which could not be captured by the used segmentation algorithm.

One of the limitations of this work is the lack of a large ABUS dataset to test the repeatability of the proposed segmentation tool. ABUS is a relatively new technique and its clinical use is still not widely spread. Thus, there is lack of a good public dataset. Furthermore, inter and intra-reader variability is a well known issue in ultrasound imaging that can influence the performance evaluation of the automatic segmentation frameworks. Considering all these issues, the analysis presented here may not be fully extrapolated to clinical practice in its current state. However, this work presents an important step towards the development of an automatic tool for breast lesion quantification in ABUS.

Despite the developments in ABUS for screening dense breasts, one of the major limitations is the non-availability of large public ABUS dataset, which hinders the use of emerging deep learning methods based on Convolutional Neural Network (CNN), as it requires large amount of data for training the CNNs. For instance, one of the recent attempts to use deep learning approaches in ABUS volumes was made by Wang et al. [127], where authors used a 3D U-Net as the CNN architecture, and fine-tune the hyper-parameters of a pre-trained network [128]. More recently, Chiang et al. [129] designed a three stage method for tumor detection in ABUS, using a sliding window detector to detect volumes of interest.

In contrast to 3D ultrasound, mammography is widely used for breast cancer screening, especially in women with non-dense breast. Some of the mammography datasets are publicly available, in which the most popular are DDSM [47], CBIS-DDSM [48] and INbreast [49] datasets. The DDSM contains Screen-Film Mammogram (SFM) from $\approx$2500 cases, while INbreast consists of Full-Field Digital Mammogram (FFDM) from 115 cases. In the following chapter, deep learning methods are used to develop automatic CAD system for lesion detection using the two available public mammography datasets.

Chapter 4

Patch-Based Lesion Detection on Mammograms Using Deep Learning Methods

In the previous chapter, a semi-automatic lesion detection framework was presented for 3D Automated Breast Ultrasound (ABUS) volumes, and was used for temporal analysis of breast volumes. This was the first step towards the development of a semi automatic tool for breast lesion quantification in ABUS volumes. Following on from the works in previous chapter, the focus is shifted on 2D mammography images (using public datasets). This is done owing to facilitate the development of automated Computer Aided Detection (CAD) systems based on emerging deep learning technology. In this chapter, a particular class of deep learning i.e. Convolutional Neural Network (CNN) is used to develop automatic CAD system for lesion detection using the two available public mammography datasets.

4.1 Introduction

Since the past decade, research in breast image analysis has mainly focused on the development of CAD systems in order to assist radiologists in the diagnosis. Traditionally, mammographic CAD systems relied on hand-engineered features, which showed limited accuracy in complex scenarios. More recently, with

the advent of deep learning methods, CAD systems learn automatically which image features are more relevant to be used to perform a diagnosis, boosting the performance of these systems.

One of the limitations of deep learning is that it requires large training datasets, and in the field of medical imaging compiling such a large image database to extract relevant features from different diseases can be a tedious task. Thus, while deep learning strategies based on CNN are promising in terms of diagnosis of a number of diseases, there are still significant constraints due to limited training data that must be overcome. To this effect, the researchers have devised some strategies, such as (i) use image patches as inputs rather than full sized image [130–132], (ii) to expand the data artificially by using data augmentation, (iii) to use deep models trained over large databases of natural images and then fine-tune the network with the medical images (referred as transfer learning) [133, 134].

Transfer learning (also known as domain adaptation) is considered to be an efficient methodology, in which the knowledge from one image domain can be transferred to another image domain. Here, the aim is to fine-tune a pre-trained model (trained on a larger database) on a smaller dataset [135]. Azizpour et al. [136] suggested that the success of any transfer learning approach highly depends on the extent of similarity between the databases on which a CNN is pre-trained and the database to which the image features are transferred. Tajbakhsh et al. [137] debated if the use of pre-trained deep CNNs with sufficient fine-tuning could eliminate the need for training a deep CNN from scratch. The authors also analysed the influence of the choice of the training samples on the performance of CNNs, and concluded that there is a no set rule to say if a shallow tuning or deep tuning is beneficial and that optimal method is dependent on the type of application.

In this chapter, an automated framework is proposed which uses image patches for training and testing the CNNs. The concept of transfer learning is used to perform a domain adaptation between the images of different characteristics i.e. natural images, Screen-Film Mammogram (SFM) and Full-Field Digital Mammogram (FFDM). The training of CNNs is based on the image patches extracted from the SFM and FFDM.

4.2 Convolution Neural Networks: CNNs

Convolutional Neural Network (CNN) is a deep learning algorithm especially designed to work with two dimensional image data. It is capable to assign weightings to different image features in order to differentiate one from another.

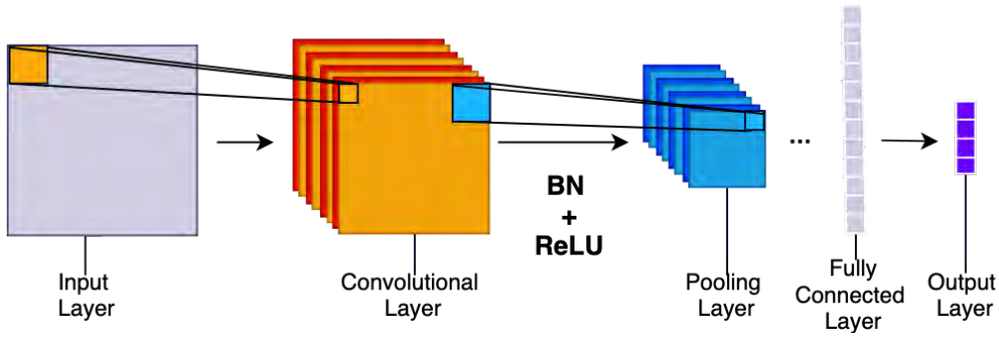


Figure 4.1: A simple CNN architecture with 1-convolutional layer is shown. The convolutional block can be repeated multiple times to increase the depth of the network.

Figure 4.1 shows a simple CNN architecture. The CNN consists of an input and output layer, multiple convolutional layers, batch normalisation (BN) layer, rectified linear unit (ReLU) or activation function, pooling layers and Fully Connected (FC) layer.

- **Input** layer contains the raw pixel values of the image along with the colour information.
- **Convolutional** layer performs an operation called convolution. It is a linear operation involving multiplication of a set of weights (filters or kernels) with the input, resulting in a two dimensional output array called feature map.
- **BN** layer normalises the output of the previous activation layer by subtracting the batch mean and dividing by batch standard deviation.
- **ReLU** is a piecewise linear function that outputs the input directly if it is positive otherwise outputs zero.
- **Pooling** layer performs a downsampling operation along the spatial dimensions (width, height). Two common functions used in pooling operation are average pooling and maximum pooling.
- **FC** layer have full connections to all activations in the previous layers.
- The last FC layer is called the **Output** layer which computes the class scores.

In this thesis, three widely used CNN architectures (VGG16, ResNet50 and InceptionV3) are evaluated. These CNNs have already proven to be excellent for image classification using the ImageNet dataset.

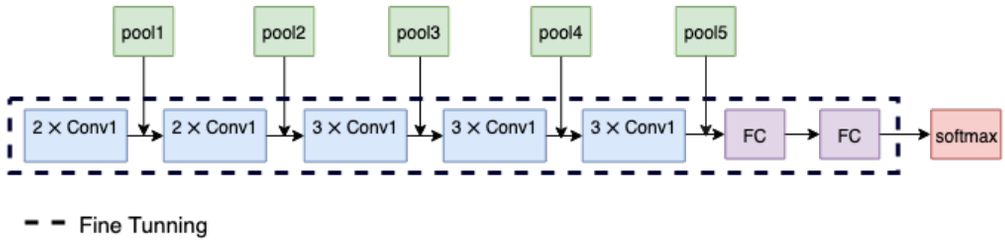


Figure 4.2: VGG16 CNN architecture showing the convolution, max-pooling, FC dense layers. Dashed line represents the fine-tuned layers.

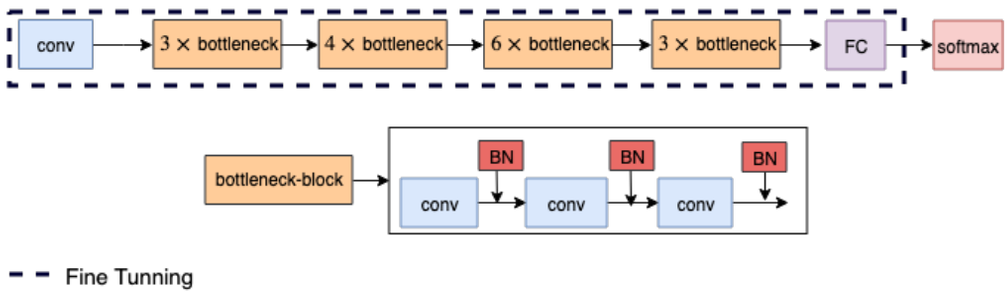


Figure 4.3: ResNet50 CNN showing the overall architecture and the bottleneck block. Dashed line represents the fine-tuned layers.

4.2.0.1 VGG

The VGG [138] network is the contribution from the Visual Geometry Group, University of Oxford, United Kingdom and consists of very small convolutional filters (3×3) with a depth of 16-19 weight layers, resulting in a simple architecture. The potential applications of the VGG based CNN model for breast cancer diagnosis have been shown in [139–144].

Herein, the VGG16 CNN (shown in Fig. 4.2) is used. The VGG16 consists of 13 convolutional layers and 2 fully connected or dense layers, followed by an output dense layer with a softmax activation function. There are also 5 max pool layers in the network. The maximum allowed input size to the VGG16 network is 224×224, and the total number of trainable parameters are 138,357,544.

4.2.0.2 ResNet50

The residual network (ResNet) [145] architecture consists of convolutional layers, pooling layers and multiple residual layers, each containing several bottleneck blocks: a stack of three convolutional layers followed by batch normalisation (BN) layers. Fig. 4.3 shows the ResNet50 structure consisting of 4 residual layers each comprising of 3, 4, 6 and 3 bottleneck blocks from bottom to top, followed by a dense layer and the output layer with softmax activation function. In total there are

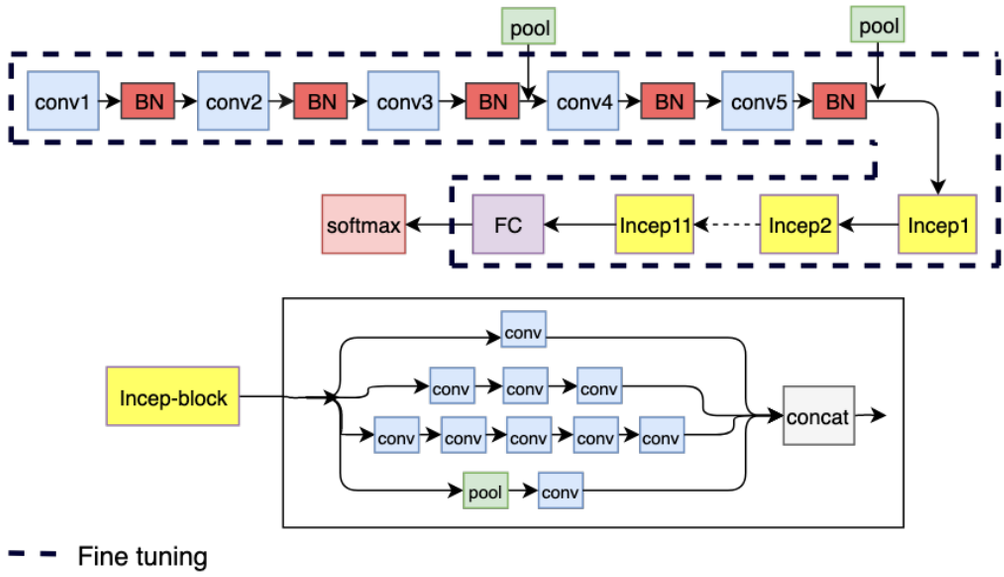


Figure 4.4: InceptionV3 CNN showing the overall architecture and the inception module. Dashed line represents the fine-tuned layers.

179 layers in ResNet50 architecture. In literature, the ResNet50 CNN model has been particularly used for breast cancer diagnosis in [142, 144, 146, 147].

4.2.0.3 InceptionV3

The InceptionV3 [148] model has been developed by Google and is also known as GoogleNet. The computational cost and memory requirement of Inception network is much lower than VGG and ResNet50, which makes it a prominent network to be used in big data scenarios. Inception network consists of a collection of Inception modules, each of which uses sets of 3x3 kernels to represent larger kernels in a computationally efficient manner. The network implemented here has 5 convolutional layers each followed by a BN layer, 2 pooling layers and 11 inception modules as shown in Fig. 4.4. The InceptionV3 CNN model has been used for breast cancer diagnosis in [72, 99, 149–151].
A summary of the three CNN networks used in this chapter is shown in Table 4.1.

4.3 Methodology

The proposed fully automated framework for mass detection is depicted in Fig. 4.5. The developed framework is initialised by extracting small regions of the image (referred to as patches) for training the CNN. The model obtained after the CNN training is firstly used to classify the unseen testing patches as mass and non-mass patches (with different probabilities). The patches are then recombined

Table 4.1: Summary of CNN architectures: including input size, number of layers, and total number of trainable parameters.

Network	Minimum Input size	Maximum Input size	No. of layers	Year	Trainable parameters
VGG16	32 × 32	224 × 224	16	2013	138,357,544
Resnet50	32 × 32	224 × 224	179	2015	25,636,712
InceptionV3	75 × 75	299 × 299	159	2016	23,851,784

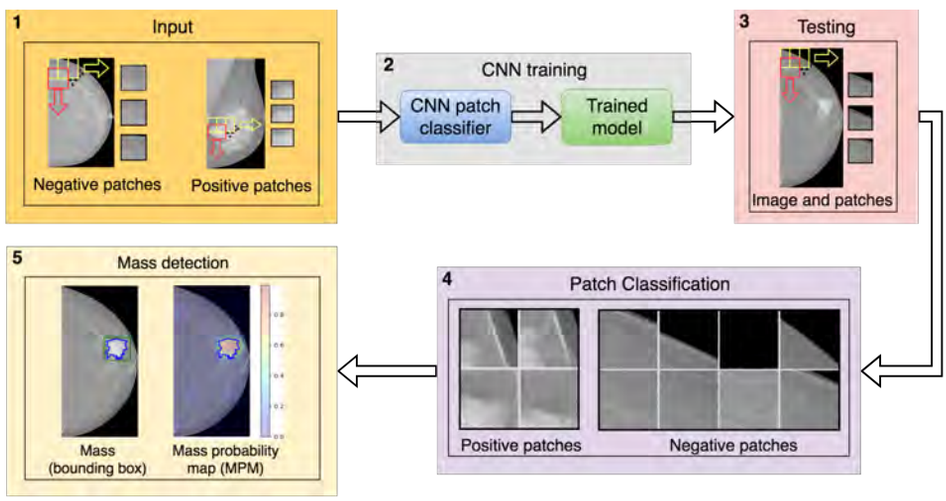


Figure 4.5: The proposed framework for automated detection of masses in mammograms. The first block shows the patch extraction strategy using sliding window for negative (non-mass candidates) and positive (mass candidates) mammograms, followed by CNN training in block 2 to obtain a trained model. The third block shows the patch extraction from a test image followed by patch classification using the trained model (shown in block 4). The block 5 shows the MPM and the detection on the original image (with the green bounding box).

to reconstruct the whole mammogram and subsequently use the classification probabilities for each patch to obtain the Mass Probability Map (MPM) for the mammogram. Finally, the probable mass region is defined by a bounding box. The complete automated framework is described in following sections.

4.3.1 Input Patch Extraction

A sliding window approach is used to scan the whole breast and extract all the possible patches from the image (see Fig. 4.5 Step 1). The total number of patches generated is controlled by the stride ($s \times s$) in both vertical and horizontal direction, which also defines the minimum overlap between the two consecutive patches. All the patches are then classified based on the annotations provided in the dataset. For example, a patch is labelled as positive (mass candidate) if the central pixel of the

patch lies inside the mass, otherwise it is assigned as a negative (no mass) label.

Since all images of the CBIS-DDSM dataset contains a lesion, an equal number of positive and negative patches are extracted from mass images. All the positive patches are firstly extracted from the annotated ROIs and then an equal number of negative patches are randomly selected from the normal region of the breast (excluding the border area patches due to high contrast difference). This provides a balanced dataset for the purpose of training. The INbreast dataset contains mammograms with and without masses, so positive patches (Pos) are extracted only from the mammograms with masses. In order to maintain a balance between positive and negative samples for the CNN training, the negative patches (Neg) are extracted from the mammograms without masses using the following formulation:

$$P_{Neg} = ceil\left(\frac{n}{N}\right), \tag{4.1}$$

where n is the number of positive patches, N is the total number of non-mass mammograms in the training or validation set, and P_{Neg} is the required number of patches to be randomly selected from each of the non-mass mammograms.

4.3.2 CNN Training

The CNNs described in section 4.2 are initially trained on the ImageNet dataset with input dimensions $224 \times 224 \times 3$, where the three dimensions represent red, green and blue colour channels. Since extracted patches from mammograms contains only one channel (grey level), each patch ($224 \times 224 \times 1$) has been replicated onto the three-colour channels to make the input patches compatible with the input of the pre-trained CNNs as done in other works [142, 146]. Normalisation also called zero centring is a standard step in medical image classification. In this work, Global Contrast Normalisation (GCN) is used. It computes the mean of intensities for each image patch, and subtracts it from each pixel of the image [142].

For CNN training, the dataset is split into training, validation and test sets. The training set is used to train the network and update its weights, while the validation set is used to measure how well the trained model is performing after each epoch. An epoch here describes the number of times the algorithm processes the entire dataset. Further, data augmentation is used to generate more samples from already existing training data. In this chapter, the negative and positive patches are augmented on-the-fly using horizontal flipping, rotation of up to 30°, and re-scaling by a factor chosen between 0.75 and 1.25, as commonly used in the literature [31, 95, 137, 142]. The optimizer used is Adam [152] and the batch size is 128 (for a

GPU of 12 GB). The cross-entropy loss function is used as the validation loss, which measures the performance of a classification model whose output is a probability value between 0 and 1. For a binary classification problem (number of classes is 2), cross entropy is calculated as

$$L = -(ylog(p) + (1-y)log(1-p)) \tag{4.2}$$

where L is the loss, y is the class label (0 or 1) and p is the predicted probability. Early-stopping can be used as regularisation technique in order to avoid over-fitting during the training. Herein, early-stopping is used as the monitor on the validation loss, such that if the loss does not improve for 10 epochs, the training process is stopped.

Firstly, the performance of different CNNs are analysed for classifying mass and non-mass regions in CBIS-DDSM dataset. For the purpose, two different training approaches are compared. Firstly, the CNNs are trained from scratch using the random weight initialization. The training is done for 100 epochs (maximum) using a learning rate of 10^{-3}. Further, the extent of transfer learning is analysed by transferring the domain from natural images to SFM. This is carried out by using the pre-trained ImageNet weights to initialize the CNNs and fine-tune all the layers of the CNN (without freezing any layer) for 100 epochs (maximum) using a learning rate of 10^{-6}. A higher learning rate is used while training the models initialized using random weights owing to the fact that training the CNN from scratch would need more time to learn the features pertaining to the images being analysed. In contrast, when the CNN is initialized using pre-trained weights (where the model has already been trained on millions of images) the features learned during initial training are sensitive to the extent of training, so a smaller learning rate is used to preserve pre-trained features while fine-tuning.

4.3.3 Mass Detection

The best performing CNN model is subsequently fine-tuned to transfer the feature domain from SFMs in CBIS-DDSM to FFDMs in the INbreast dataset. After fine-tuning the CNNs using the INbreast training and validation dataset (using a learning rate of 10^{-6}), mass detection is performed in a fully automated manner without any human intervention. This is achieved using the following steps (see blocks 3-5 in Fig. 4.5):

1. Firstly all the possible patches are extracted from each image using sliding window approach.

2. The patches are analysed using the trained CNN to obtain the mass probability of each patch. The image is then reconstructed by stacking the patches in sequence from left to right and top to bottom (similar to a patch extraction using sliding window approach), with the stride value ($s \times s$) defining the overlap between the patches. The MPM is then generated using the linear interpolation of the mass probabilities (on each patch) as

$$Mass\ probability = \frac{\sum mass\ probability\ of\ overlaping\ patches}{number\ of\ overlaping\ patches} \tag{4.3}$$

3. The MPM is then thresholded at different probability levels. This step results in the creation of different regions (each region represents a probable mass) in the mammogram such that each pixel in these regions have the probability greater than the chosen threshold value.

4. A bounding box is created to enclose each probable region using connected component analysis. A mass is considered detected if the Intersection over Union (IoU) between the bounding box and the annotated ground truth is greater than 0.2, as suggested in earlier works [31, 90, 153, 154].

4.3.4 Evaluation Metrics

The evaluation metrics used in this chapter are: (a) the testing accuracy of the model, (b) Area Under Receiver Operating Curve (AUROC), (c) Free-Response Operating Curve (FROC) curve, and (d) Intersection over Union (IoU). The FROC curve is used to evaluate the performance of the detection tool on the INbreast dataset and is plotted between the fraction of correctly identified lesions as True Positive Rate (TPR) and the number of False Positives per Image (FPI) for all decision thresholds. The TPR is evaluated as $\mu \pm \sigma$ where μ and σ refer to the mean and standard deviation.

IoU is an evaluation metric used to measure the accuracy of an object detector on a particular dataset. Any algorithm that provides predicted bounding boxes as output can be evaluated using IoU. The IoU between the detection and ground truth bounding box is computed as

$$IoU = \frac{\text{Area of overlap}}{\text{Area of union}}. \tag{4.4}$$

The detection is considered as a False Positives (FP), True Positives (TP) or False

Negatives (FN) if the IoU$<$ 0.2, IoU$>$ 0.2 or IoU$=$ 0 respectively, as depicted in Fig. 4.6.

4.4 Experimental Results

This section presents the different training and transfer learning experiments performed in order to evaluate the CNN models. Firstly, a description of the processing applied to the datasets used in this chapter is presented in section 4.4.1. Secondly, the domain of convolutional features are transferred from natural images to SFM, i.e. section 4.4.2, and is achieved by training the CNNs on the CBIS-DDSM dataset. Later, in section 4.4.3 the transfer learning is performed between SFM and FFDM, and compared with the results obtained when transferring the CNN domains directly from natural images to FFDM. Lastly, in section 4.4.4 the framework proposed in section 4.3.3 for detection masses in whole mammogram is used to analyse the INbreast dataset.

4.4.1 Processing of Datasets

In this chapter, the CBIS-DDSM [48] and INbreast [49] datasets are used. The descriptions of these datasets have been provided in chapter 2. Herein, details regarding the processing of these datasets specific to this chapter are provided.

The **CBIS-DDSM** dataset contains pixel-wise annotations for the Region of Interest (ROI), e.g. masses, calcifications, as well as lesion's pathology i.e benign or malignant from 1,597 cases. The CBIS-DDSM dataset is composed of SFM images which implies a non-homogeneous intensity distribution in the background as shown in Fig. 4.7a. Therefore, Otsu segmentation [66] is used to segment the breast region and the background region (as shown in Fig. 4.7b). The mammograms are then cropped to the breast profile as shown in Fig. 4.7c.

The **INbreast** dataset is composed of FFDMs acquired from 115 cases leading to a total of 410 images. From these, a total of 116 masses can be found in 107 mammograms from 50 cases. Herein, the cases with follow-up studies are not considered (different acquisition times) as different cases, and thus resulting in a total of 108 cases. The dataset contains pixel-level mass annotations and histological information about the type of cancers. The dataset also contains some mammograms with multiple masses. It is found that in four mammograms the lesions are very close and their bounding-boxes overlap, so these lesions are as one single lesion. Thus the total number of masses is considered to be 112 instead of

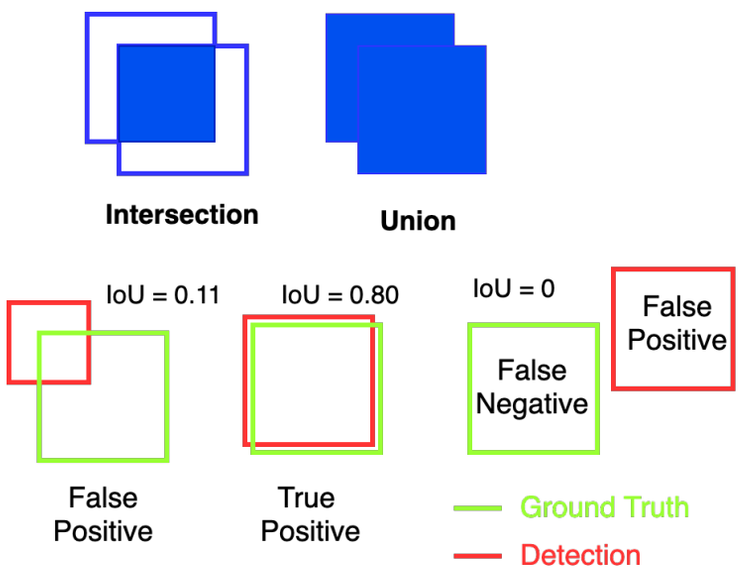


Figure 4.6: Schematic for the computation of IoU, and representation of FP, TP and FN in terms of IoU (used in this thesis).

116. Regarding pre-processing of these FFDMs, global thresholding is performed to segment the breast region from the background (see Fig. 4.8) and all right breasts are mirrored horizontally in order to keep the left orientation of all mammograms. Note that in all cases the original resolution of the processed DICOM mammograms is used. In both the datasets, the patches of size 224×224 pixels are generated using a stride of 56×56 pixels, and are used as the input to the CNNs.

4.4.2 Transfer Learning from Natural Images to SFM

In this experiment a domain adaption of the CNN is performed from the natural images to mammography dataset. For the purpose, a CNN model pre-trained using a large database of natural images (i.e. ImageNet) is fine-tuned on a dataset of SFM (i.e. CBIS-DDSM). The standard training and test splits is used as provided in the CBIS-DDSM dataset. The stride value is selected to obtain a trade-off between the computational requirements and the number of training samples. Table 4.2 provides the details of the number of patches ($\approx$65,000) extracted from the CBIS-DDSM dataset.

The three CNNs previously described, i.e. VGG16, ResNet50 and InceptionV3 pre-trained using ImageNet are fine-tuned on the training dataset of CBIS-DDSM. The classification results are then compared against those obtained when the CNNs are randomly initialized and used for the purpose of classifying masses. To demonstrate the potential of transfer learning for lesion classification, the CNN training was repeated five times (owing to the randomness of training process).

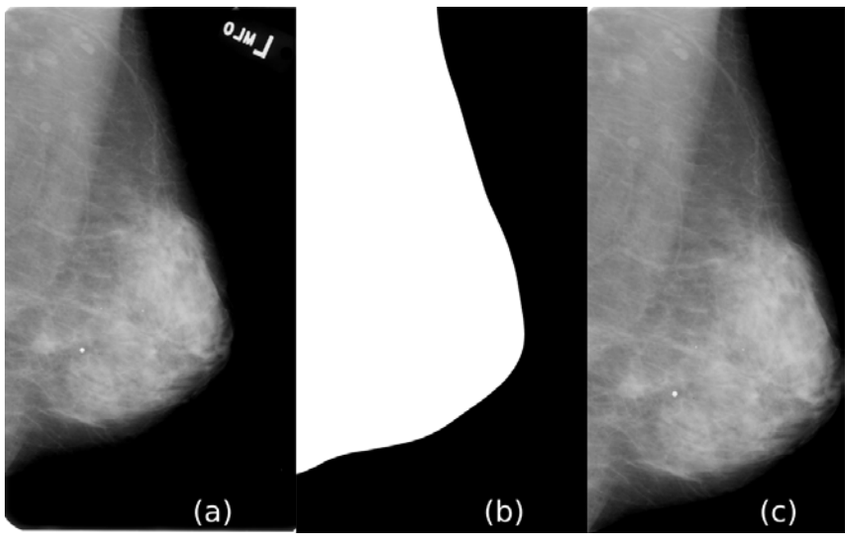


Figure 4.7: Sample mammogram from CBIS-DDSM, (a) original, (b) segmented, and (c) mammogram cropped to breast profile.

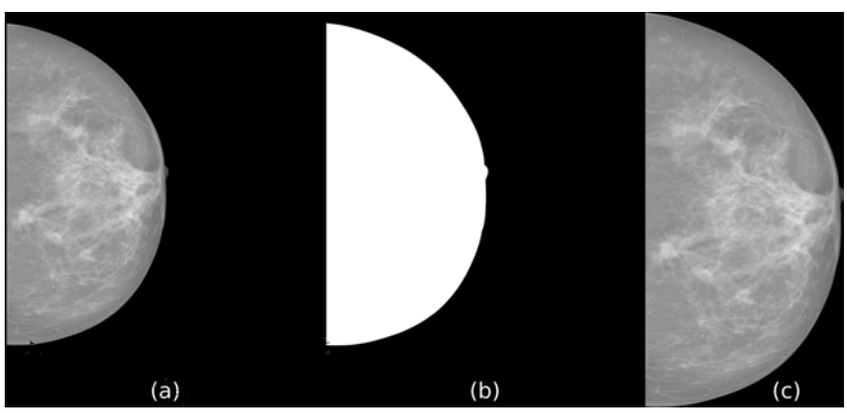


Figure 4.8: Sample mammogram from INbreast, (a) original, (b) segmented, and (c) mammogram cropped to breast profile.

Table 4.3 compares the testing results between the random and ImageNet weight initialization. Note that in all cases, the initialisation with ImageNet weights obtained a better accuracy compared to random initialization, and InceptionV3 CNN obtained the highest testing accuracy $84.16\% \pm 0.19$ and AUROC of 0.93 ± 0.01. Moreover, as shown in Fig. 4.9, the randomly initialized CNN required a large number of epochs to converge than the pre-trained InceptionV3 demonstrating the benefits of pre-training on ImageNet.

The obtained results showed that the difference in performance (testing accuracy) of the pre-trained InceptionV3 with pre-trained ResNet50 and VGG16 respectively is statistically significant ($p \ll 0.01$). Also, for each CNN, the difference in performance between the random and ImageNet initialization is found to be statistically significant ($p \ll 0.01$). For the rest of the chapter, all experiments

Table 4.2: Description of CBIS-DDSM dataset used for training the CNN

Dataset	Splits	Cases		Images		Patches		Stride
		Pos	Neg	Pos (mass)	Neg	Pos	Neg	
CBIS-DDSM	Train	553	-	985 (1055)	-	25979	25979	56
	Validation	138	-	246 (263)	-	6210	6210	56
	Test	201	-	361 (378)	-	8694	8694	56

Table 4.3: Mass classification performance (testing accuracy) for mass and non-mass region in CBIS-DDSM dataset for VGG16, ResNet50 and InceptionV3 where μ and σ refer to the mean and standard deviation for five independent training results.

Model	Pre-trained	Time per Epoch (s)	Testing accuracy ($\mu \pm \sigma$)	AUROC
VGG16	No	518	$82.39\% \pm 0.52$	0.90 ± 0.01
	Yes	465	$83.69\% \pm 0.24$	0.92 ± 0.01
Resnet50	No	483	$82.30\% \pm 0.70$	0.91 ± 0.01
	Yes	438	$83.69\% \pm 0.15$	0.92 ± 0.01
InceptionV3	No	338	$82.10\% \pm 0.58$	0.90 ± 0.01
	Yes	310	$84.16\% \pm 0.19$	0.93 ± 0.01

are performed using the pre-trained InceptionV3 CNN model which provides the best mass classification results on the CBIS-DDSM dataset.

4.4.3 Transfer Learning from SFM to FFDM

Since both the INbreast and CBIS-DDSM are mammography datasets, with the only difference being the mode of acquisition (SFM and FFDMs), the feature space of the CNN for one is assumed to be very likely to be relevant to the other dataset. So, in this experiment the performance of transfer learning is analysed between the images of similar domains. A 5-fold cross validation is used to analyse the performance on the whole dataset. The dataset is divided into training (60%), validation (20%) and testing (20%) sets on the case level per fold. The distribution is performed in an stratified manner to ascertain equal ratios of normal and abnormal cases. Table 4.4 provides the details of the distribution of images into training, validation and test set (along with the number of patches) in the INbreast dataset.

The best model obtained from section 4.4.2 (CBIS-DDSM pre-trained InceptionV3 model) is fine-tuned (using 10^{-6} learning rate) on the INbreast dataset. Table 4.5 shows the impact of transfer learning on InceptionV3 CNN. The results indicate that using the transfer learning between the images of similar

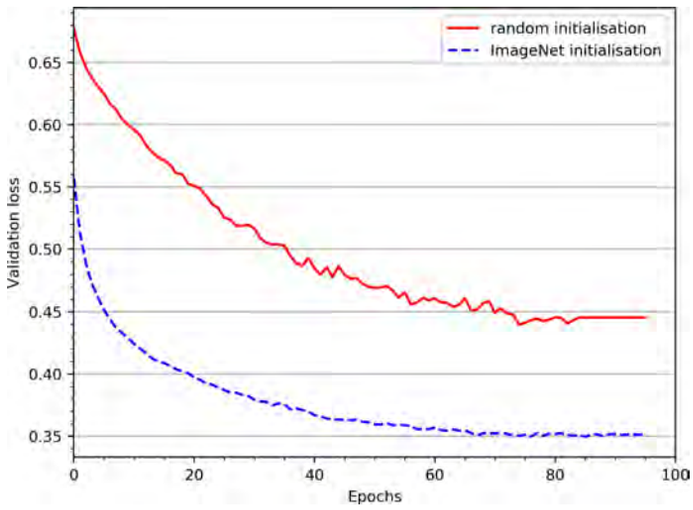


Figure 4.9: Validation loss for random and Imagenet initialisation of InceptionV3.

Table 4.4: Description of INbreast dataset used for training the CNN

Dataset	Splits	Cases		Images		Patches		Stride
		Pos	Neg	Pos (mass)	Neg	Pos	Neg	
INbreast	Train	30	35	66 (68)	191	2020	2101	56
	Validation	10	11	20 (21)	61	539	549	56
	Test	10	12	21 (23)	51	882	918	56

domains, the testing accuracy improves from $85.29\% \pm 4.29$ (ImageNet→INbreast) to $88.86\% \pm 2.96$ (ImageNet→CBIS-DDSM→INbreast).

4.4.4 Automated Mass Detection in FFDM

In this section, the training cascade: ImageNet→CBIS-DDSM→INbreast with InceptionV3 is used to automatically detect masses in the test set FFDM without any human intervention. Here, mass detection is performed on the INbreast dataset using a 5-fold cross validation strategy. The full mammogram from the test set of each fold is divided into small patches using a sliding window approach with stride of 56×56, and the trained CNN model (per fold) is then used to classify these patches into mass and non-mass regions and generate the MPM images (see Fig. 4.5 block 5). The mass detection is then performed following the methodology described in Step 3 and 4 of section 4.3.3.

The detection performance on the full INbreast dataset is analysed using FROC curves as shown in Fig. 4.10, where the upper and lower bounds are presented in 95% confidence interval. Note that ImageNet→INbreast configuration is also

Table 4.5: Testing accuracy for classifying mass and non-mass region in INbreast dataset, where μ and σ refer to the mean and standard deviation for five fold cross validation.

Model	Pre-trained weights	Training cascade	Testing accuracy
InceptionV3	ImageNet	ImageNet → INbreast	$85.29\% \pm 4.29$
InceptionV3	CBIS-DDSM	ImageNet → CBIS-DDSM → INbreast	$88.86\% \pm 2.96$

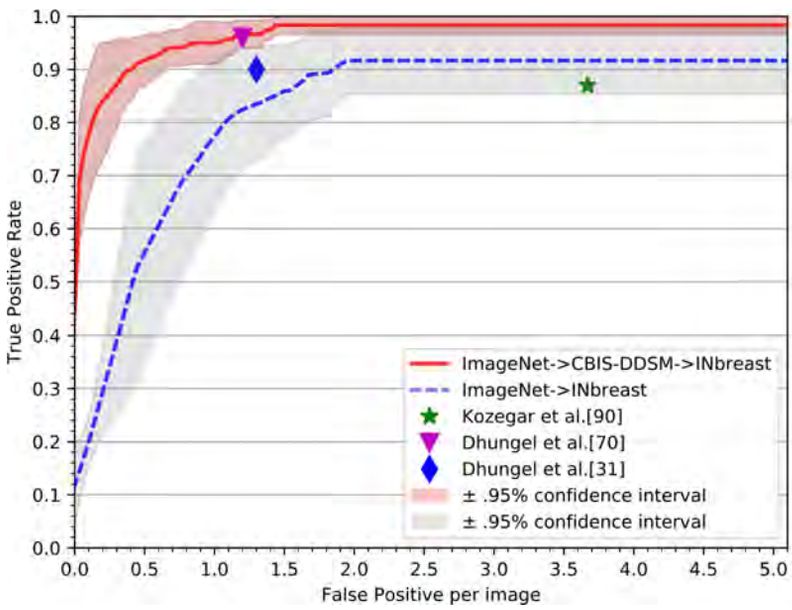


Figure 4.10: FROC curve for mass detection on INbreast dataset: testing performance of InceptionV3 pre-trained on CBIS-DDSM and fine-tuned on INbreast are plotted using 5-fold cross-validation strategy. The operating points from the literature are shown for direct comparison with the proposed framework.

included in Fig. 4.10 for the purpose of comparison. It is observed that for the same evaluation measure of IoU ≥ 0.2, the performance of CNN is substantially higher when the transfer learning is performed between the images of similar domains (i.e. ImageNet → CBIS-DDSM → INbreast) with TPR of 0.98 ± 0.02 at 1.67 FPI, compared to that obtained when using database of natural images (ImageNet→INbreast) with TPR of 0.91 ± 0.07 at 2.1 FPI. Results reported in the literature for the same task using INbreast dataset are also included in Fig. 4.10 for further comparison.

To analyse the performance across different mass sizes, the lesions are divided in three categories following radiological criteria, i.e. small lesions of area<1 cm^2 (25 images), medium size lesions ranging 1 cm^2 $<$area<4 cm^2 (42 images), and large lesions with area>4 cm^2 (38 images), and analysed the performance of the proposed detection framework for each case. Fig. 4.11 shows that the small lesions have a TPR of 0.89 at 0.5 FPI, while the medium and large lesions have the same

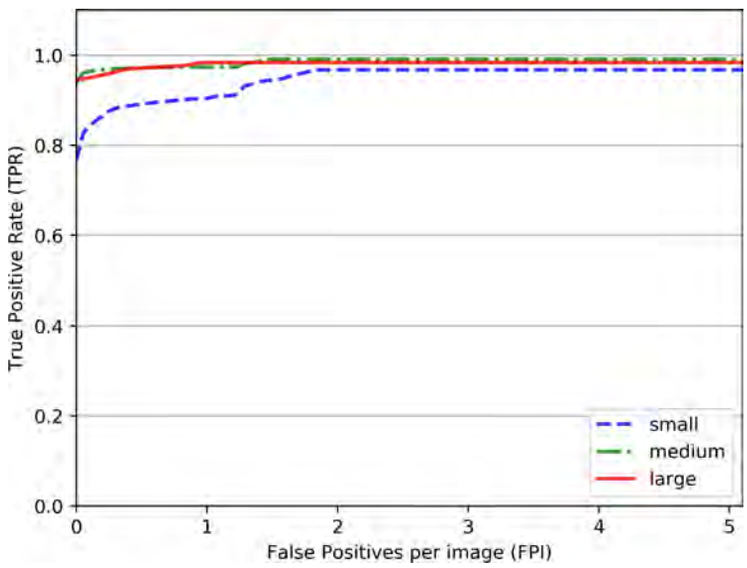


Figure 4.11: FROC curve showing the performance of the proposed framework on INbreast dataset for different lesion sizes (ImageNet → CBIS-DDSM → INbreast).

TPR of 0.97 at 0.5 FPI. Consequently, the detection performance is inferior for small lesions below 1 cm^2, which are more challenging to detect.

4.4.5 Qualitative Analysis

Fig. 4.12 illustrates examples of mass detection on few testing images (unseen during training) performed using the best model obtained from CBIS-DDSM → INbreast fine-tuning. Fig. 4.12 (a-h) show examples of correctly detected masses in Cranio-Caudal (CC) and Medio-Lateral Oblique (MLO) views with variable lesion size and lesion contrast. In addition, Fig. 4.12 (i, j) show examples of FP detections (red squares), where dense tissue area and a nodule in the pectoral region mimic the appearance of lesion like structure. Note that the proposed method is unable to detect only 2 masses (very small size) out of the total of 112 lesions within the INbreast dataset, and these are shown in Fig. 4.12 (k, l).

4.5 Discussions and Conclusions

In this chapter, an end-to-end mass detection framework was presented using the CNN-based patch classification approach. To generalize the applicability of the proposed framework, three different CNN architectures were analysed using two public datasets containing SFMs and FFDMs.

The interesting aspect of the transfer learning is to reuse the CNN model

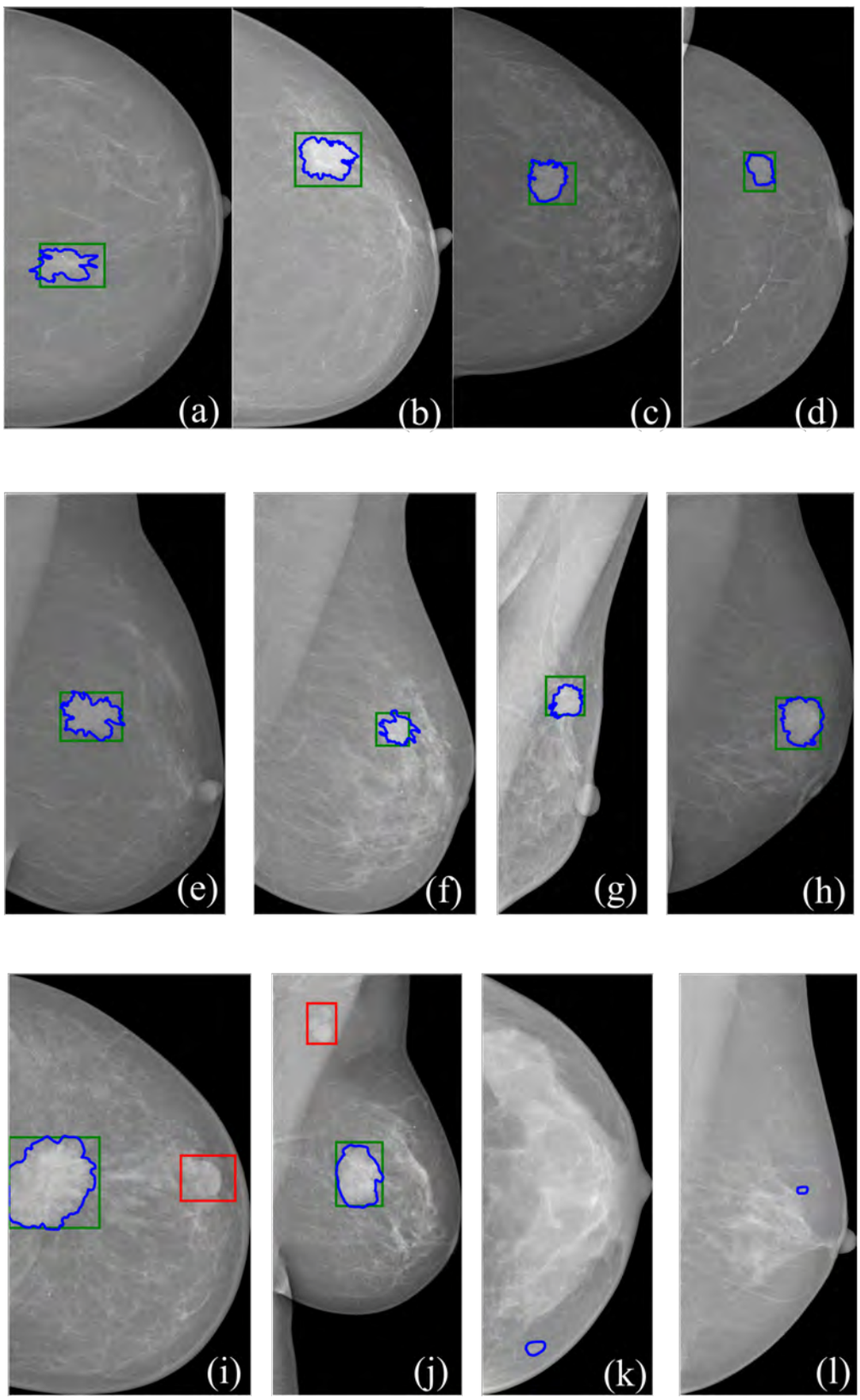


Figure 4.12: Mass detection examples from INbreast using the ImageNet → CBIS-DDSM → INbreast strategy. (a-h) illustrate correct detections, (i, j) represent FP cases and (k, l) correspond to missed mass cases. Blue contours represent the GT (masses), green bounding boxes correspond to the automated detection (TP), and red squares show missing lesions (FP).

pre-trained for a completely different problem, and improve the accuracy of a given task using less complex algorithms and also the training time is shorter. In this regard, firstly the benefit of transfer learning was analysed between two entirely different image domains, i.e. natural images and mammograms (SFM). In this context, a comparison was done between the performance of CNNs with randomly initialized weights vs. pre-trained (ImageNet) weight initialization for the purpose of mass classification in SFM, showing higher performances for the pre-trained models.

As shown in Table 4.3, it can be seen that despite the differences in two image domains, the pre-trained CNNs performed substantially better than the randomly initialized CNNs. These results gave confidence on the applicability of transfer learning in the context of mammograms. This is supported by the fact that the pre-trained CNN is able to efficiently use the information of universal features and patterns learned from the ImageNet. In section 4.4.3, transfer learning was done between SFM and FFDM, and compared with the results obtained by transferring the CNN domains from natural images to FFDM. The results in Table 4.5 indicated that substantially higher classification accuracy can be obtained by performing transfer learning between the images of similar domains i.e. SFM$\rightarrow$FFDM.

In section 4.4.4, the proposed detection framework produced the best TPR of 0.98 ± 0.02 at 1.67 FPI and a TPR of 0.92 ± 0.04 at 0.5 FPI with IoU of 0.2. A higher detection threshold (IoU $\geq$ 0.5) was also analysed and resulted in TPR of 0.82 ± 0.2 at 1.7 FPI, compared to TPR of 0.98 ± 0.02 for IoU $\geq$ 0.2. Thus, the FPI is dominated by the number of negative images; which are $\sim$4 times larger than the positive images in the INbreast dataset, so changing the IoU does not have a large effect on the FPIs. The detection performance of the proposed framework was superior in terms of TPR when compared with other state-of-the-art methods using the INbreast dataset (see Table 4.6 and Fig. 4.10) on various other operating points.

For the purpose of pre-processing two different approaches were investigated: (i) the image intensities were scaled between 0-255 before extracting the patches, and (ii) GCN normalisation was applied to obtain the zero mean over the input patches. Both approaches showed different impact on the fine-tuning process, with the GCN approach showing higher performance compared to the scaling approach. Thus, the results in section 4.4.2 and 4.4.3 were performed using GCN pre-processing.

There are some important things to note about training the CNN: (i) the CNNs were also fine-tuned by training only the last few layers (also referred as shallow tuning) as discussed in the literature [137, 142], with no significant improvement

Table 4.6: Comparison between proposed framework and results published in the literature using INbreast dataset, where μ and σ refer to the mean and standard deviation for five fold cross validation.

Methods	**TPR ($\mu \pm \sigma$) at FPI (IoU)**	**# Images (INbreast)**
Kozegar et al. [90]	0.87 at 3.67 (0.2)	107
Dhungel et al. [70]	0.96 ± 0.03 at 1.2 (0.2)	410
Dhungel et al. [31]	0.90 ± 0.02 at 1.3 (0.5)	410
Proposed framework	0.87 ± 0.07 at 0.25 (0.2) 0.90 ± 0.06 at 0.44 (0.2) 0.95 ± 0.04 at 0.79 (0.2) 0.98 ± 0.02 **at 1.67 (0.2)**	410

in the classification results on the CBIS-DDSM and INbreast dataset [155]. Finally fine-tuned the CNNs by training all the layers at a small learning rate; (ii) It was also observed that the random weight initialization took a larger number of epochs to converge than initializing using ImageNet weights (see Fig. 4.9).

It is to be noted that the mass detection framework proposed in this chapter used small regions of the mammograms (patches) for training the CNN. These patches were extracted using a sliding window approach and required a selection of appropriate stride value which is used to slide a fixed size patch window to map the full image. The network predicts the probability on each patch, which was then used to obtain the MPM for the whole mammogram (see Fig.4.5). To analyse the performance of network with respect to the stride used, varying patch strides were analysed (for testing), which results in smaller or larger number of patches than those presented in section 4.4. Increasing the stride also increases the similarity in the input data (owing to higher overlap), and vice-versa. This step demanded a trade off between the accuracy and the computational cost (for stride of 56×56 the average detection time per image was $\approx 30 - 50$ seconds depending on size of the image). Very large strides resulted in a poor localized predictions, whereas very small strides required very high computational cost ($\approx 70 - 100$ seconds for stride of 28×28).

An alternative approach to overcome the high computational cost associated with selection of smaller strides (which results is very large number of patches) is to use an efficient methodology to select the most desirable candidates (patches) to compute the probabilities or predictions, and subsequently use them for the detection process. In the following chapter, a mass detection framework will be presented which uses the full mammogram as the input (instead of patches) and

is based on the “recognition using regions” paradigm [156], which has been very successful in the field of object detection in natural images [157].

Chapter 5

Lesion Detection in Whole Mammogram using Faster R-CNN

In the previous chapter, a patch-based breast lesion detection framework using Convolutional Neural Network (CNN) was presented to detect masses in publicly available mammogram datasets. In this chapter, a lesion detection framework is presented which uses the whole mammogram for the training and testing of CNN. For the purpose, Faster R-CNN model is used which has InceptionV2 [148] as the base CNN model, and is pre-trained on the MS-COCO dataset [158]. This selection is done based on a trade-off between the speed and accuracy. In the final phase of research, subset of a large scale mammographic dataset OMI-DB was collected in collaboration with Royal Surrey County Hospital (UK). This database was not accessible during the work done in chapter 4, and has been used for the training and testing of Faster R-CNN model presented in following sections.

5.1 Faster R-CNN

The Region-Based CNN (R-CNN) are usually characterised by deep CNN architectures with an enhanced capability to learn more complex features than the shallow ones. The R-CNN was proposed originally by Ross Girshick [159] in 2014, and achieves excellent object detection accuracy by using a deep convolutional network to perform object detection. The R-CNN is a three-stage process: (1)

region proposal generation, (2) feature extraction using CNN, and (3) object classification and localisation. In spite of advancements over traditional methods, the training process in the R-CNN is slow, as the features are extracted for each object proposal without sharing the computation.

In R-CNN the region proposals are generated using a selective search method [157] as an alternative to the exhaustive search approach to capture the object location. In selective search, firstly the image is over-segmented based on the intensity of the pixels using the fast segmentation method of Felzenszwalb and Huttenlocher [160]. Thereafter, bounding boxes are generated corresponding to each segments and added to the region proposals. Secondly, a similarity factor is calculated between neighbouring regions, and two most similar regions are grouped together. The process of grouping the most similar regions is repeated until the whole image becomes one region. At each step, larger segments are formed and added to the list of region proposals. Hence the region proposals are created from smaller regions to larger regions in a bottom-up approach. This methodology efficiently reduces the search space for the object detection task, but also results in a large number of redundant proposals.

Later in 2015, Girshick proposed an improved R-CNN method referred to as Fast-RCNN [161]. In Fast-RCNN, the input image is directly fed to the CNN to generate a convolutional feature map. From the convolutional feature map, the region of proposals are identified and extracted using the ROI pooling layer and fed into the fully connected layer. Thus this method was faster compared to R-CNN, as it did not require to feed large number of region proposals to CNN every time, instead the convolution operation is done only once per image to extract the feature map. A multi-task loss function based on training classification accuracy and bounding box regression was introduced. It resulted in saving the additional expense on storage space, and improved both accuracy and efficiency with more reasonable training schemes.

Faster R-CNN [79] is the modified version of Fast-RCNN. The main difference between the two approaches is that Fast-RCNN uses computationally expensive selective search method to generate the regions of interest, while Faster R-CNN uses Region Proposal Network (RPN) for this purpose, which is much faster compared to the selective search approach. In this work, the implemented version of the Faster R-CNN within the Tensorflow [162] object detection Application Programming Interface (API) [163] is used. The Faster R-CNN has been widely used for detecting objects particularly in natural image datasets, for example, PASCAL VOC [164], MS-COCO [158] etc. In this API [163] a collection of pre-trained detection models

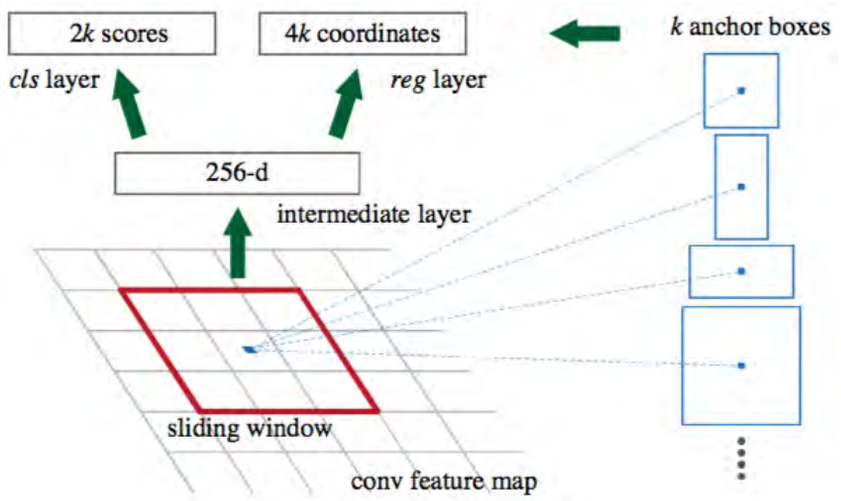


Figure 5.1: RPN in Faster R-CNN [79], showing the anchors obtained at different scales and aspect ratios in the intermediate layers.

is provided which includes: 21 models pre-trained on MS-COCO dataset [158], 1 model on KITTI dataset [165], 6 models on Open Images [166] and 2 models on iNaturalist species [167]. In the field of medical imaging, the use of object detection techniques has been limited because of limited availability of labelled datasets.

5.2 Methodology

As the first step, an image (mammogram) is used as the input and then forwarded through the CNN (InceptionV2 [148]) to produce feature map of the image. A RPN is then created on the top of extracted features of the CNN, which is then trained to detect and localize objects on the image. The architecture of RPN is shown in Fig. 5.1. The CNN model takes as input the entire image and produces the conv feature maps. A window of size $n \times n$ slides over the feature maps and outputs a features vector linked to two Fully Connected (FC) layers i.e. box-regression layer (reg) and box-classification layer (cls). K number of rectangular boxes (called as anchors) are convoluted with each sliding window.

At this step, anchors or bounding boxes are created using different scales and aspect ratios to detect objects with varying shapes and sizes. The anchors are given an objectness score about how good they are in terms of enclosing a lesion on the mammogram. Now, the highest scoring anchors (different sizes) are passed to the second stage of the network, where a classification and regression problem is solved to accurately detect the presence of lesions and at the same time refine the coordinates of the anchors to precisely detect the lesions. Then, the

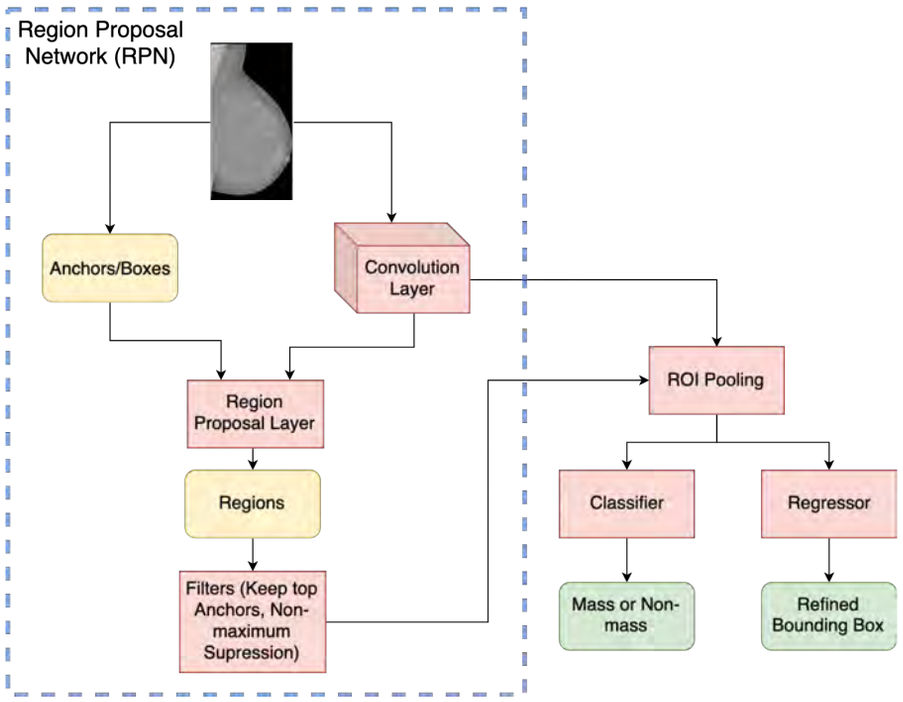


Figure 5.2: Flow chart of the Faster R-CNN, showing the RPN and overall pipeline. The convolutional layer of the InceptionV2 are used as feature extractor which shares the feature map with the RPN, generating the region proposals. The classification and regression problem are solved on these proposals to generate the final bounding box informations.

best predictions are obtained by using non-maximum suppression on the detected overlapping objects.

In this chapter, the Faster R-CNN model proposed by Ren et al. [79] is adapted to generate region proposals for varying shapes and sizes, and are labelled as positives (representing masses) and negatives (representing background or non-mass region) for all the mammograms. More precisely, as it has been implemented in Tensorflow, the input to the model is the Tfrecords of training and validation data (containing the mammograms along with class definitions, bounding box coordinates etc.). Tfrecord is an efficient way of storing the data in binary format, and is very helpful when dealing with large datasets. Binary data take less disk space, are faster to copy, and can be read efficiently and fastly through the disk, speeding up the training process. The flowchart of the overall methodology is shown in Fig. 5.2.

5.2.1 Training and Hyperparameter Tuning

In Faster R-CNN, the training is performed using the hyperparameter “keep aspect ratio” with the maximum height and width of the mammograms in the entire dataset.

Here, the model training is performed using stochastic gradient descent optimizer with momentum with a batch size of 1 (as model is trained with different size mammograms). The learning rate is heuristically decreased in steps after every 25,000 iterations in the configuration file.

The anchors are created with a base size of 256 pixels, three aspect ratios (0.5, 1.0 and 2.0) and five different scales (0.1, 0.2, 0.5, 1.0 and 2.0). The different aspect rations and scales are empirically selected. A total of 15 anchors are generated at a defined pixel location. At the second stage, the detections are processed with non-max separation using an Intersection over Union (IoU) threshold of 0.05. This is done to avoid overlapping detection boxes, as in mammograms it is less likely to have overlapping masses.

5.2.2 Evaluation Metric

In this work, the objectness score (measure of closeness of the detected object to a certain class of object) is used to evaluate the performance, so for the assessment of classification and detection framework only the bounding boxes with confidence probability greater than a particular threshold are considered. Herein, the confidence threshold is varied between 0.01-0.99 to plot the Free-Response Operating Curve (FROC) curve. For the detection framework, a lesion is considered to be properly detected (a True Positives (TP)) if the IoU defining the overlapping area between the predicted box and the Ground-Truth (GT) box is greater than a certain threshold. The higher the IoU, the better the predicted location of the box for a given object.

The qualitative assessment is made using the confusion matrix to compute the sensitivity, specificity and the accuracy of the classification framework as

$$sensitivity = \frac{TP}{TP+FN}, \quad specificity = \frac{TN}{TN+FP} \tag{5.1}$$

where, TP, TN, FP and FN are the true positives, true negatives, false positives and false negatives per mammograms respectively. Also the Area Under Receiver Operating Curve (AUROC) is used to evaluate mass classification results. The accuracy of the detection problem is assessed using the FROC curve which is plotted as the function of sensitivity versus the False Positives per Image (FPI).

Table 5.1: Data description of OMI-DB: Pos refers to positives (masses) and Neg refers to negatives (non-masses). Hologic and Philips scanners have two different resolutions.

Scanners		Hologic (OPT-H)	GE (OPT-G)	Siemens (OPT-S)	Philips (OPT-P)
Resolution μm		70, 60	100	70	50, 83
Cases	Pos	2042	103	32	242
	Neg	842	104	-	-
Images	Pos	3856	195	63	455
	Neg	3478	406	-	-
Masses	Benign	390	9	2	19
	Malign	2294	112	48	397
	Unknown	502	38	0	33
other abnormalities	Benign	95	2	2	5
	Malign	754	31	11	47
	Unknown	182	14	0	5

5.3 Experimental Results

5.3.1 Dataset: OMI-DB

In chapter 2, an overview of the OMI-DB [22] has been provided. In this chapter, the images acquired from the scanners of different manufacturer in the OMI-DB dataset (Table 5.1) are used for evaluating the proposed mass detection framework. In Table 5.1, it can be seen that there are a very few cases from the Siemens (only 32), which are not enough for training the network and thus these cases are not used in this thesis. Further, as per the author's knowledge, Philips has stopped to commercialise their scanners, so images from Philips are also not investigated. In this chapter, the mammograms obtained from the two manufacturer: Hologic (OPT-H) and GE (OPT-G) are used for the training and testing of the CNN.

There are several abnormalities in the dataset as masses, calcifications, architectural distortions, focal asymmetries, and combination of masses with calcifications, distortions or asymmetries, or masses with architectural distortion or masses with focal asymmetry. As an illustration, some abnormalities in the dataset are shown in Fig. 5.3. Since, the aim of this thesis is to focus on the detection of

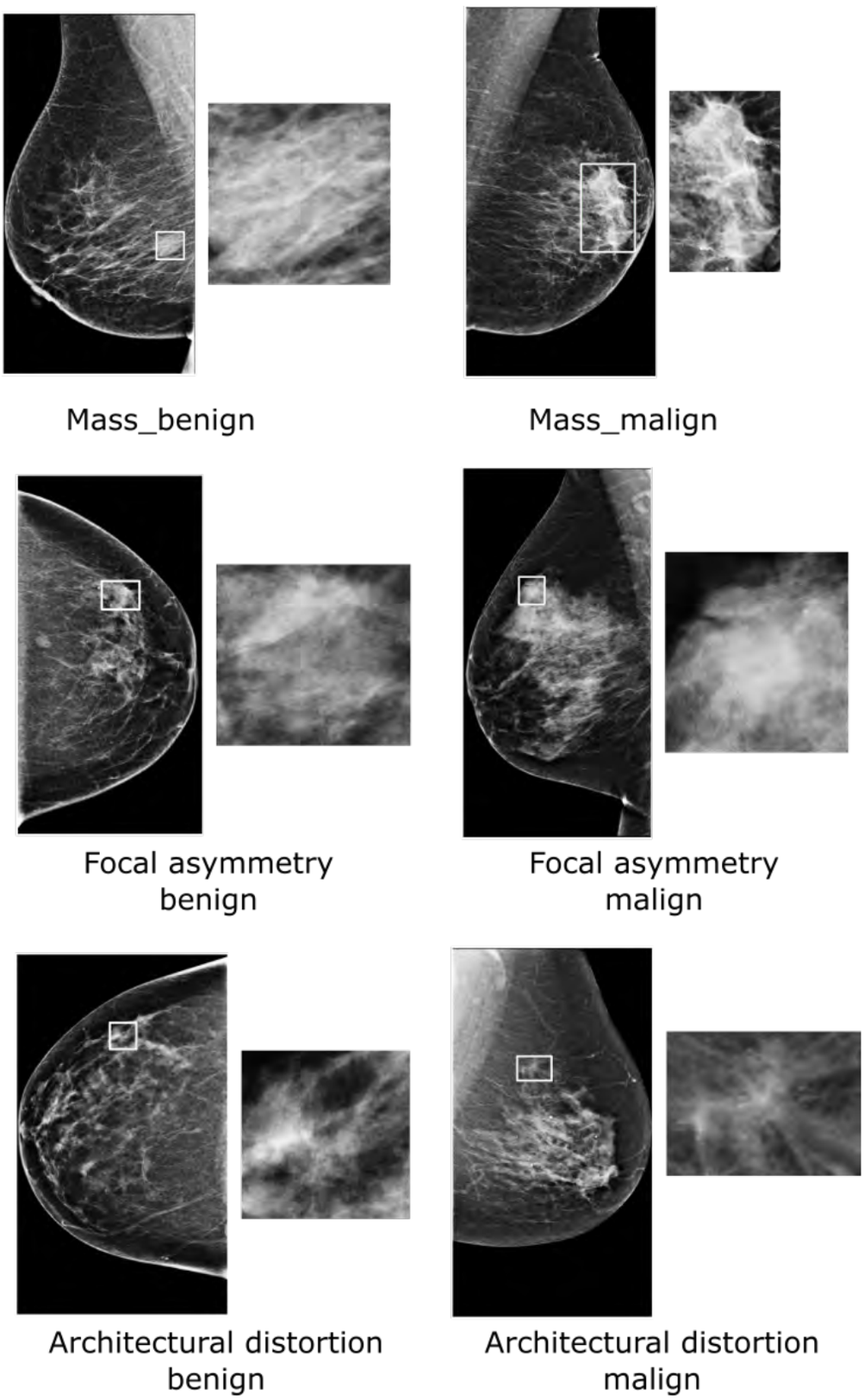


Figure 5.3: Sample mammograms with abnormalities in the OMI-DB dataset (Hologic scanner).

masses (or mass like abnormality) in mammograms, so the mammograms with only calcifications are not considered in any case as positives. Furthermore, the dataset contained some cases with implants, and several mammograms show non-standard views i.e. magnified view of the abnormality [168], these cases are also not considered. The architectural distortion and focal asymmetry are together referred as other abnormalities in the later text.

The masses are categorised into benign and malignant based on the BI-RADS ratings listed in the dataset. Note that the BI-RADS ratings are obtained after biopsy, confirming the nature of abnormality. The details of complete dataset is shown in Table 5.1. In the dataset, some masses are present whose Breast Imaging Reporting and Data System (BI-RADS) ratings are not available (referred as unknown in the text), so these cases are only used for the training, and not for the testing purpose. The OPT-H and OPT-G respectively contained 2042 and 103 positive cases with abnormalities in either one of the views of mammograms, 842 and 104 normal cases i.e. without any abnormalities. In addition, the dataset contained mammograms of different sizes and resolutions, so firstly the mammograms are cropped to the breast profile, and subsequently down-sampled to $200\mu m$.

5.3.2 Lesion Detection on Large Mammography Dataset

In this section, firstly the scope of transfer learning is analysed to perform a domain transfer from natural images to mammograms. This is done by adapting the Faster R-CNN model pre-trained on a large dataset of natural images (MS-COCO) to detect masses in the mammograms specially taking into account that they can be acquired from different scanners. The pre-trained Faster R-CNN model is fine-tuned using a dataset of 7,334 mammograms acquired using the Hologic scanner. The dataset consists of 3,856 positive and 3,478 normal mammograms, and are divided into training, validation and test sets in the ratio of 70%, 10% and 20% respectively. The division of images is done on case basis such that all the mammograms from an individual case belongs exclusively in either training or testing set.

The model obtained after training is evaluated on 1,344 unseen mammograms in the testing set (655 with masses). The predictions are made on each mammogram and compared against the available GT bounding box annotations. Fig. 5.4 shows a plot of the TPR versus IoU between the GT and detection results. It can be seen that TPR only reduces very slightly (0.93$\rightarrow$0.88) for 0.1$<$IoU$<=$0.3, and starts to fall sharply for IoU$>=$ 0.6. This demonstrates that the mass detection model accurately

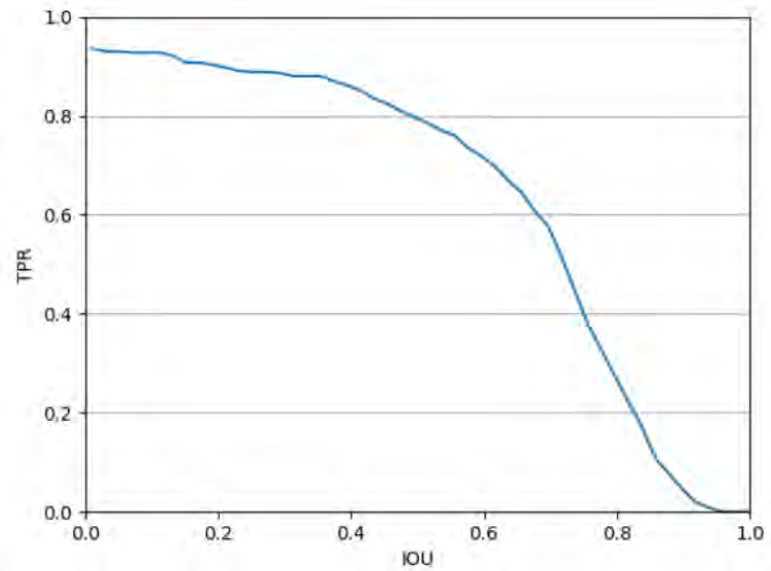


Figure 5.4: Comparison of TPR vs IoU for the detection of masses in OPT-H dataset

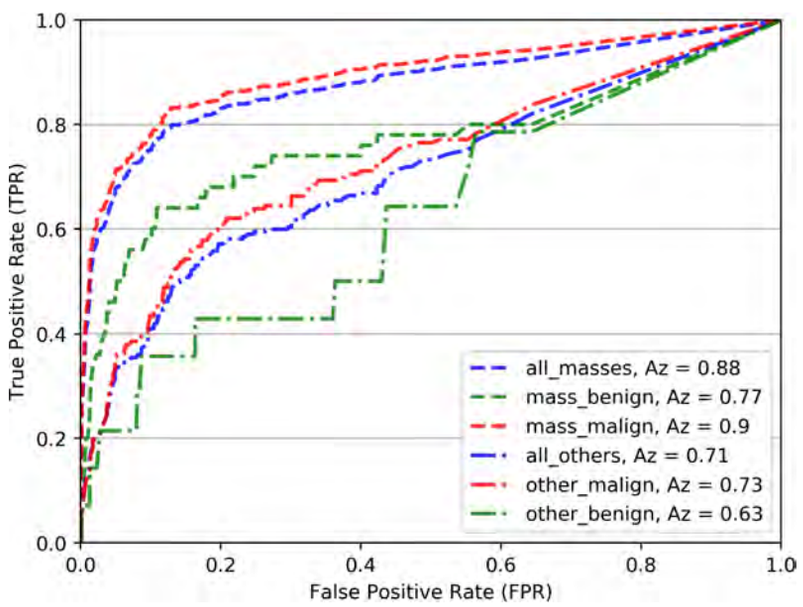


Figure 5.5: Classification performance on OPT-H dataset using ROC curve are shown for malign and benign masses and other abnormalities.

detects masses up to a value of IoU= 0.3, which is higher than the value of 0.2 which is commonly used in the current state-of-the-art mass detection models [90, 169–171]. In this chapter, an IoU threshold of 0.1 is used to compare the predicted and GT results (also used by Ribli et al. [74]). The performance of network is analysed on different abnormalities in the dataset, with the ROC curve plotted in Fig. 5.5 showing the TPR as a function of FPR. As expected the majority of lesions in the OPT-H dataset are malignant, the model's performance is better for malignant

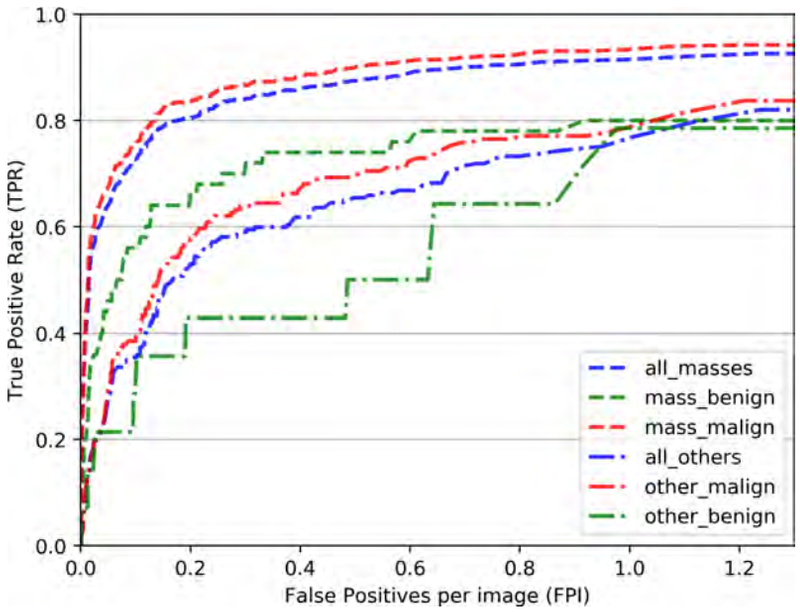


Figure 5.6: FROC curve for mass detection on OPT-H dataset shown for malign and benign masses and other abnormalities.

masses (AUROC= 0.90) compared to the benign masses (AUROC= 0.77).

In order to test the performance of the model to accurately localise the lesions, the FROC curve is plotted in Fig. 5.6 showing the sensitivity as a function of the number of false positives detected per image. As seen in Fig. 5.6, the proposed framework is able to detect malignant masses with a sensitivity of 0.94 at 1.20 FPI and 0.90 at 0.52 FPI. For the benign masses, the model results in a sensitivity of 0.80 at 0.92 FPI and 0.74 at 0.33 FPI. In addition, the model obtains a sensitivity of 0.84 at 1.20 FPI and 0.79 at 0.98 FPI for other malignant and benign abnormalities respectively. The performance of the proposed mass detection framework is summarized in Table 5.2.

5.3.3 Lesion Detection on Small Mammography Dataset

The limited availability of large annotated dataset has been a limiting factor for the success of deep learning methodologies in the field of breast cancer imaging. In this regards, here the transfer learning methodology is used to fine-tune the Faster R-CNN model pre-trained on a large mammography dataset to detect masses in small mammography datasets obtained using different scanners.

5.3.3.1 Mammograms from GE scanner

The model trained in section 5.3.2 is tested on a small dataset of 150 positive (with masses) and 406 negative mammograms (without masses) obtained using a GE

Table 5.2: Sensitivity, Specificity, AUROC results are shown separately for malign and benign abnormalities in OPT-H dataset. Note that the two operating points on the FROC curve are also shown; the first is TPR at 0.3 FPI and second is the best TPR result obtained.

Evaluation metric		sensitivity	specificity	AUROC	TPR at FPI
Masses	Malign	0.84	0.8	0.9	0.87 at 0.3, 0.94 at 1.20
	Benign	0.68	0.8	0.77	0.72 at 0.3, 0.80 at 0.92
	All	0.82	0.8	0.88	0.84 at 0.3, 0.93 at 1.23
Other abnormalities	Malign	0.6	0.8	0.73	0.59 at 0.3, 0.84 at 1.20
	Benign	0.43	0.8	0.63	0.43 at 0.3, 0.79 at 0.98
	All	0.57	0.8	0.71	0.64 at 0.3; 0.82 at 1.24

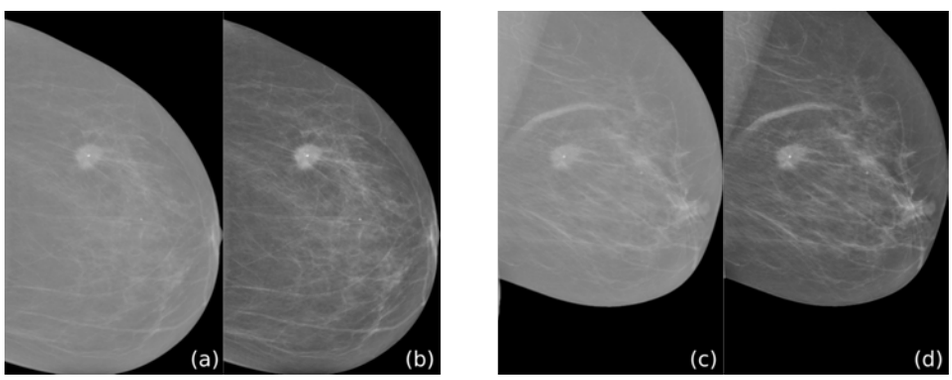


Figure 5.7: Normalization applied to images from the GE scanner, (a,c) are the original and (b,d) represent the normalised mammograms.

scanner referred as OPT-G. The images in the OPT-G dataset have very different contrast compared to the images in the OPT-H dataset, so a window width (WW) and a window center (WC) normalization are applied to the GE images [172] as shown in Fig. 5.7. The WW and WC information is obtained from the DICOM header of mammograms.

The performance of the model trained on the OPT-H dataset is firstly used directly to detect and localise masses in the mammograms in the OPT-G dataset, resulting in a sensitivity of 0.70 at 0.43 FPI and AUROC$=$ 0.77. Thereafter, fine-tuning strategy is used to further adapt the feature domain of the trained Faster R-CNN to detect masses. For this purpose, the dataset is divided based on individual cases into the training (60%), validation (20%) and test sets (20%), and a 5-fold cross-validation strategy is used to test all the mammograms in the OPT-G dataset obtaining a sensitivity of 0.91 ± 0.06 at 1.7 FPI and mean AUROC$=$ 0.87. The results are summarised in Table 5.3, and the ROC and FROC curves are shown in Fig. 5.8 and 5.9 respectively.

Table 5.3: Sensitivity, Specificity, AUROC results shown for OPT-H model tested directly on OPT-G dataset, and OPT-H model fine-tuned on OPT-G dataset. Note that the two operating points on the FROC curve are shown to establish a comparison between the performance of different trainings.

Model	Trained on OPT-H	Fine tuned on OPT-G
Sensitivity	0.70	0.85 ± 0.06
Specificity	0.73	0.73 ± 0.02
AUROC	0.77	0.87 ± 0.05
TPR at FPI	0.70 at 0.43	0.83 ± 0.07 at 0.43
		0.91 ± 0.06 at 1.69

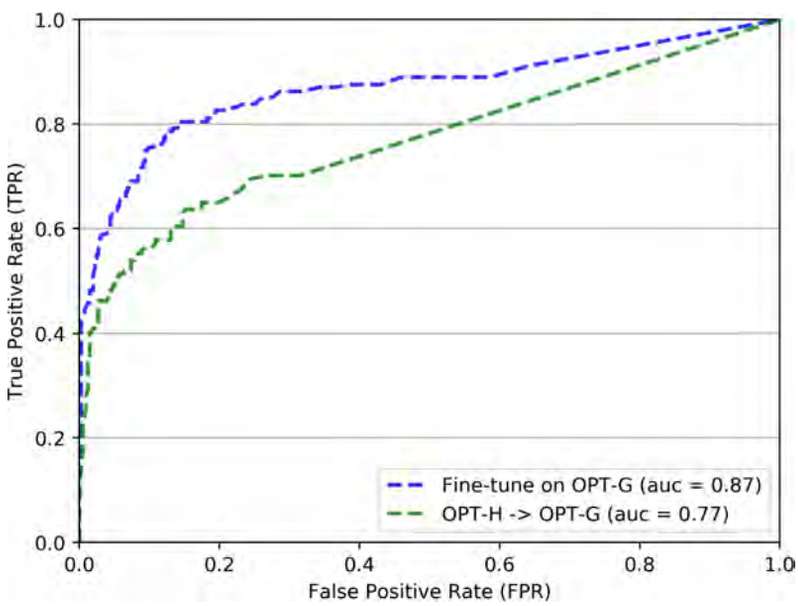


Figure 5.8: ROC curve for mass classification: OPT-H model tested directly on OPT-G dataset, and OPT-H model fine-tuned on OPT-G dataset.

5.3.3.2 Mammograms from INbreast

Dataset Processing

The INbreast dataset is composed of Full-Field Digital Mammogram (FFDM) from 108 cases, from which only 50 cases have cancer (details presented in chapter 2). Since, the mammograms in INbreast dataset have a lower contrast compared to the images in the OPT-H dataset, an adjustment is made on the windows of the pixel levels. The images are normalised (similar to Ribli et al. [74]) such that the pixel values are clipped to be minimum 500 pixel lower and maximum 800 pixels higher than the mode of the pixel value distribution (excluding the background) and were rescaled to the 0–255 range (see Fig. 5.10). The mammograms were also cropped

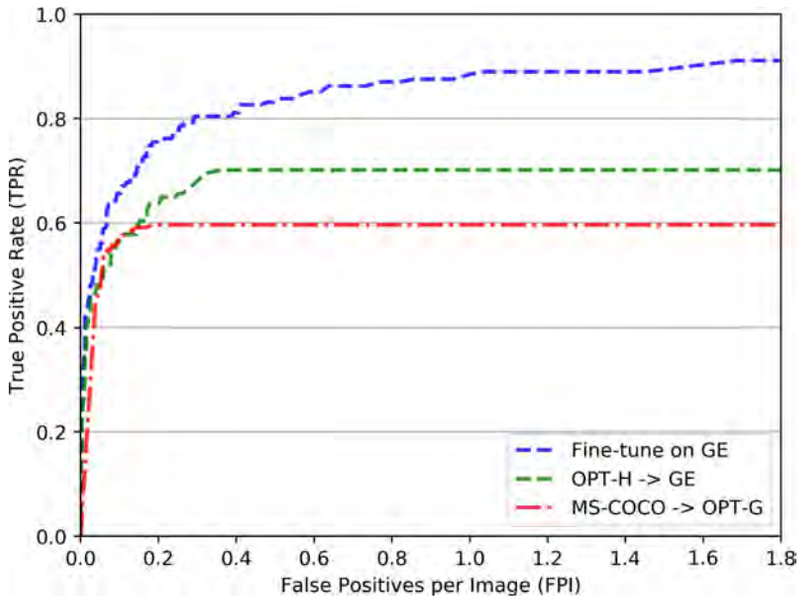


Figure 5.9: FROC curve for mass detection on OPT-G dataset: OPT-H model tested directly on OPT-G, and OPT-H model fine-tuned on OPT-G.

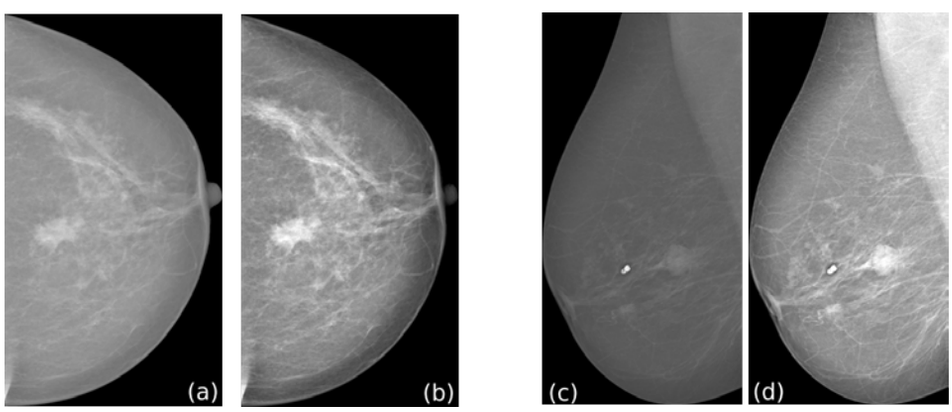


Figure 5.10: Normalization applied to INbreast, (a,c) are the original and (b,d) represent the normalised mammograms.

to the breast profile to reduce storage.

The dataset is divided into training (60%), validation (20%) and testing (20%) sets on the case level per fold. Thereafter, an augmented INbreast training dataset is created by rotating only the positive mammograms five times i.e. $0°, \pm 5°, \pm 10°$ for the purpose of training the model (as stated later in Table 5.4). A 5-fold cross validation strategy is employed to test the detection framework, so that each of the mammogram (without augmentation) is tested at least once. The distribution of the mammograms across the folds is shown in Table. 5.4.

Mass detection

In this section, firstly the performance of the model pre-trained on OPT-H dataset is

Table 5.4: Image distribution in augmented INbreast dataset for training, validation and test sets for 5-folds. Note that augmentation is only applied on the Pos training images.

		fold 0	fold 1	fold 2	fold 3	fold 4
Training	Pos	350	350	320	325	380
	Neg	183	183	185	190	176
Validation	Pos	20	25	20	31	19
	Neg	70	58	62	56	57
Test	Pos	25	20	31	19	20
	Neg	58	62	56	57	70

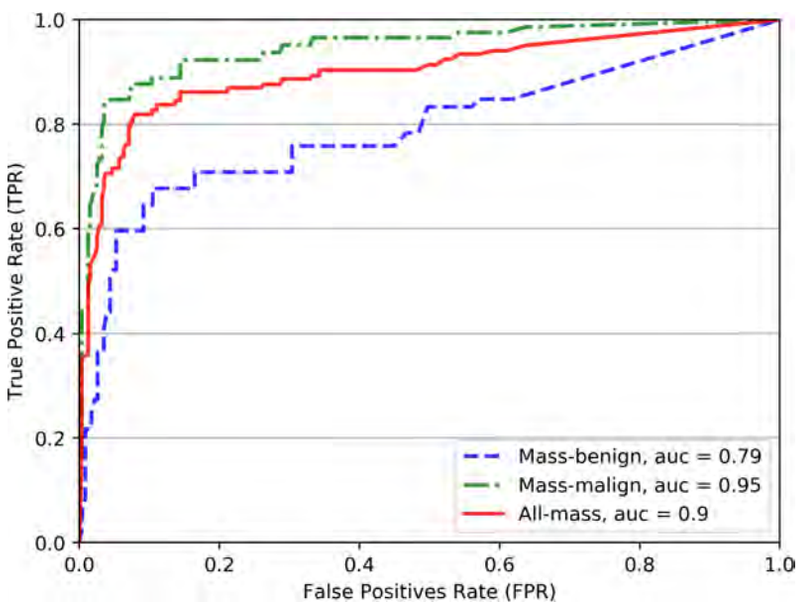


Figure 5.11: ROC curve for mass classification on INbreast dataset: Three curves are shown to compare the performances on malign, benign and all masses.

analysed to directly detect the masses in the INbreast dataset, obtaining AUROC= 0.89 for the malignant masses and AUROC= 0.67 for the benign masses. The FROC analysis resulted in a sensitivity of 0.87 at 0.32 FPI for the malignant masses and 0.55 at 0.32 FPI for the benign masses.

Secondly, the pre-trained Faster R-CNN model is fine-tuned on the augmented INbreast dataset. The detection performance is analysed using the ROC and FROC curves as shown in Fig. 5.11 and Fig. 5.12 respectively. For the malignant, benign and all masses: AUROC= 0.95, AUROC= 0.79 and AUROC= 0.90 are obtained using the ROC curve. The detection framework obtains a sensitivity of 0.99 ± 0.03 at 1.17 FPI for the malignant masses, 0.85 ± 0.08 at 1.0 FPI for the benign masses, and 0.95 ± 0.03 at 1.14 FPI for all masses. Comparing the results obtained here,

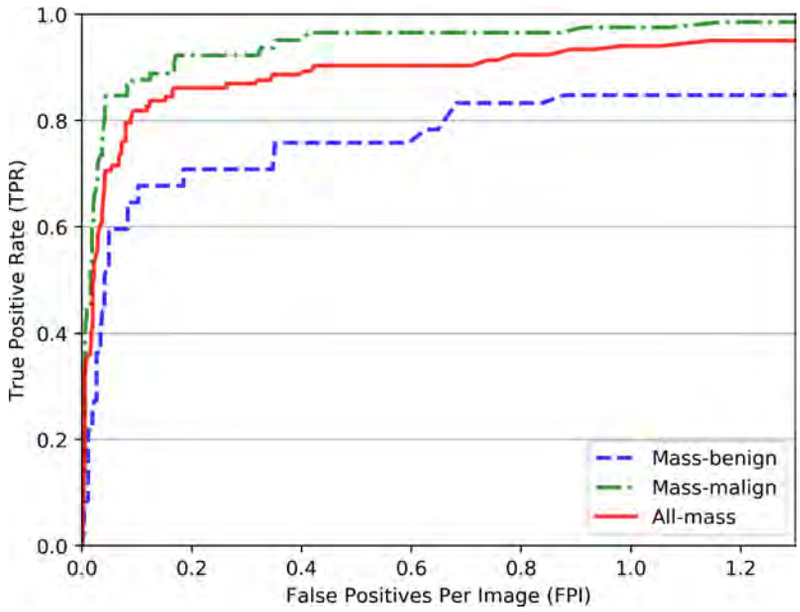


Figure 5.12: FROC curve for mass detection on INbreast dataset. Three curves are shown to compare the performances on malign, benign and all masses.

Table 5.5: Sensitivity, Specificity and AUROC for OPT-H model tested directly on INbreast dataset, and OPT-H model fine-tuned on INbreast dataset. Note that the results are shown separately for malign and benign masses, and the two operating points on the FROC curve are shown to establish a comparison between the performance of two different trainings.

Model	Trained on OPT-H		Fine-tuned on INbreast	
	Malignant	Benign	Malignant	Benign
Sensitivity	0.87	0.55	0.95 ± 0.18	0.71 ± 0.18
Specificity	0.73	0.73	0.70 ± 0.07	0.70 ± 0.08
AUROC	0.89	0.67	**0.95**	**0.79**
TPR at FPI	0.87 at 0.32	0.55 at 0.32	0.92 ± 0.08 at 0.32	0.71 ± 0.18 at 0.32
			0.99 ± 0.03 **at 1.17**	0.85 ± 0.08 **at 1.0**

with the ones presented in chapter 4, an improved performance is obtained in terms of substantial reduction in the number of false positives. The results are summarised in Table 5.5.

5.3.4 Qualitative Analysis

In Fig. 5.13 some examples of mass detection results are visualised on the mammograms in the OPT-H datasets. Several prediction results are shown: the top two rows show mammograms with precise predictions of masses; GT annotations

are displayed in green and the predicted boxes with their confidence scores are displayed in yellow. In Fig. 5.13 (i,j) FP detections are shown (in red) along with TP, Fig. 5.13 (k,l) show undetected masses.

Several mass detection results in the OPT-G and INbreast datasets are visualised in Fig. 5.14 and Fig. 5.15 respectively. Fig. 5.14 (a-h) and Fig. 5.15 (a-d) show a single mass detection result at the precise position with high confidence score in each mammogram, and Fig. 5.15 (e,h) shows detection of several masses in the same mammogram. Some of the undetected masses are shown in Fig. 5.14 (i) and Fig. 5.15(k,l).

5.4 Discussions and Conclusions

In this chapter, a lesion detection framework was presented which is based on a large scale dataset using Faster R-CNN. The model was trained and tested on a large dataset of FFDM and the lesions (or masses) were localised using the bounding box annotations. The automatic framework takes a full mammogram as the input and is able to provide the localisation of the lesion within this mammogram as the output.

The concept of transfer learning was used and a base model pre-trained for detecting objects in natural images was re-trained with the mammography dataset. For this purpose, a large private dataset of $\approx$ 7400 FFDM (OPT-H) was used to train a Faster R-CNN model (pre-trained on MS-COCO database). Mammograms in the OPT-H dataset contained both malignant and benign masses, and also normal cases (without a finding). The dataset contained the BI-RADS ratings and also the biopsy proven results. As using only the BI-RADS rating for classifying into malignant and benign might not be very accurate, so the testing results were presented on the masses which were biopsy proven to be malignant or benign. As shown in Fig. 5.5 the performance on the malignant masses (AUROC= 0.90) was substantially higher than on benign masses (AUROC= 0.77), which can be attributed to the fact that more than 70% of the total masses were malignant and thus model would be trained more efficiently to detect the malignant masses. Also, a separate analysis was presented for the detection of other abnormalities (focal asymmetry and architectural distortions).

The impact of transfer learning to adapt the Faster R-CNN model trained on a large dataset of mammograms to a smaller mammography dataset was evaluated using a small private dataset (OPT-G) and also the public dataset INbreast. In section 5.3.3.1, it was shown that images in the OPT-G had low contrast compared to the OPT-H dataset mammograms, hence normalization of the images was

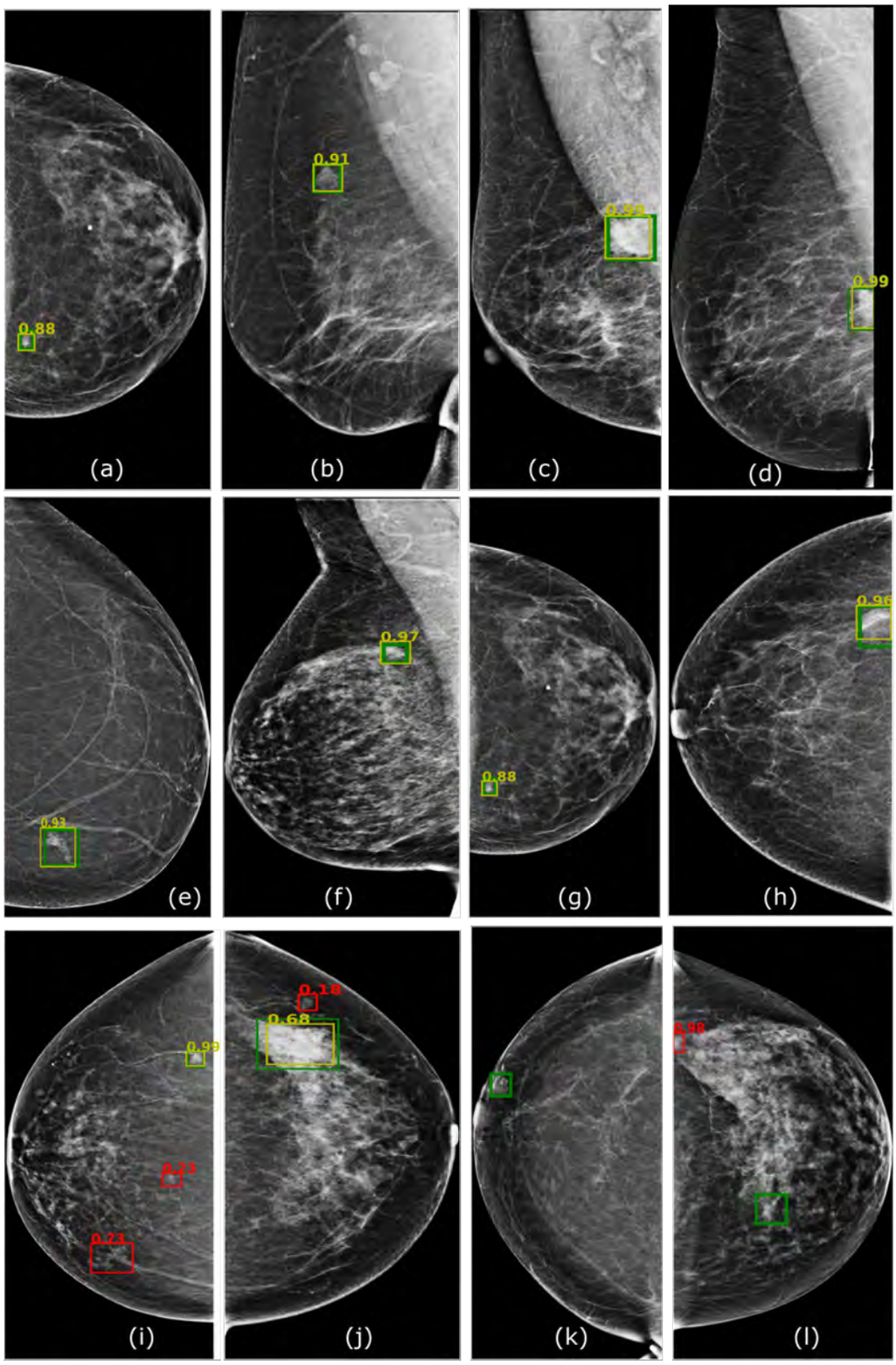


Figure 5.13: Mass detection results in OPT-H dataset, (a-h) demonstrate detections with high objectness score, (i,j) shows some detections with FPs, and (k,l) shows the undetected masses. (green: GT box, yellow and red: detection box)

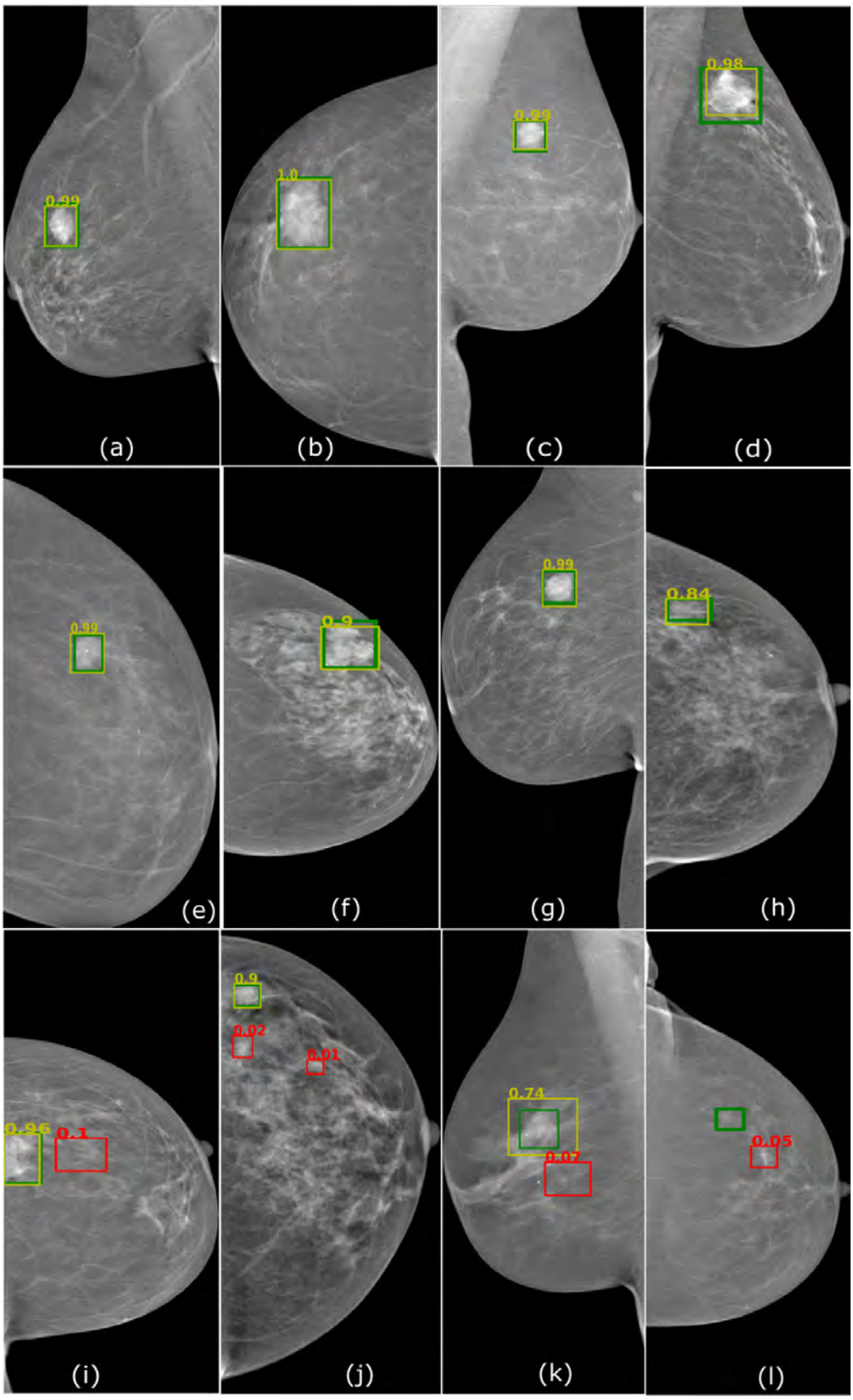


Figure 5.14: Mass detection results in OMI-G dataset, (a-h) shows detections with high confidence score, (i-k) shows detections with FPs, and (l) shows undetected mass (green: GT box, yellow and red: detection box). The numbers shown in images corresponds to the confidence of being mass.

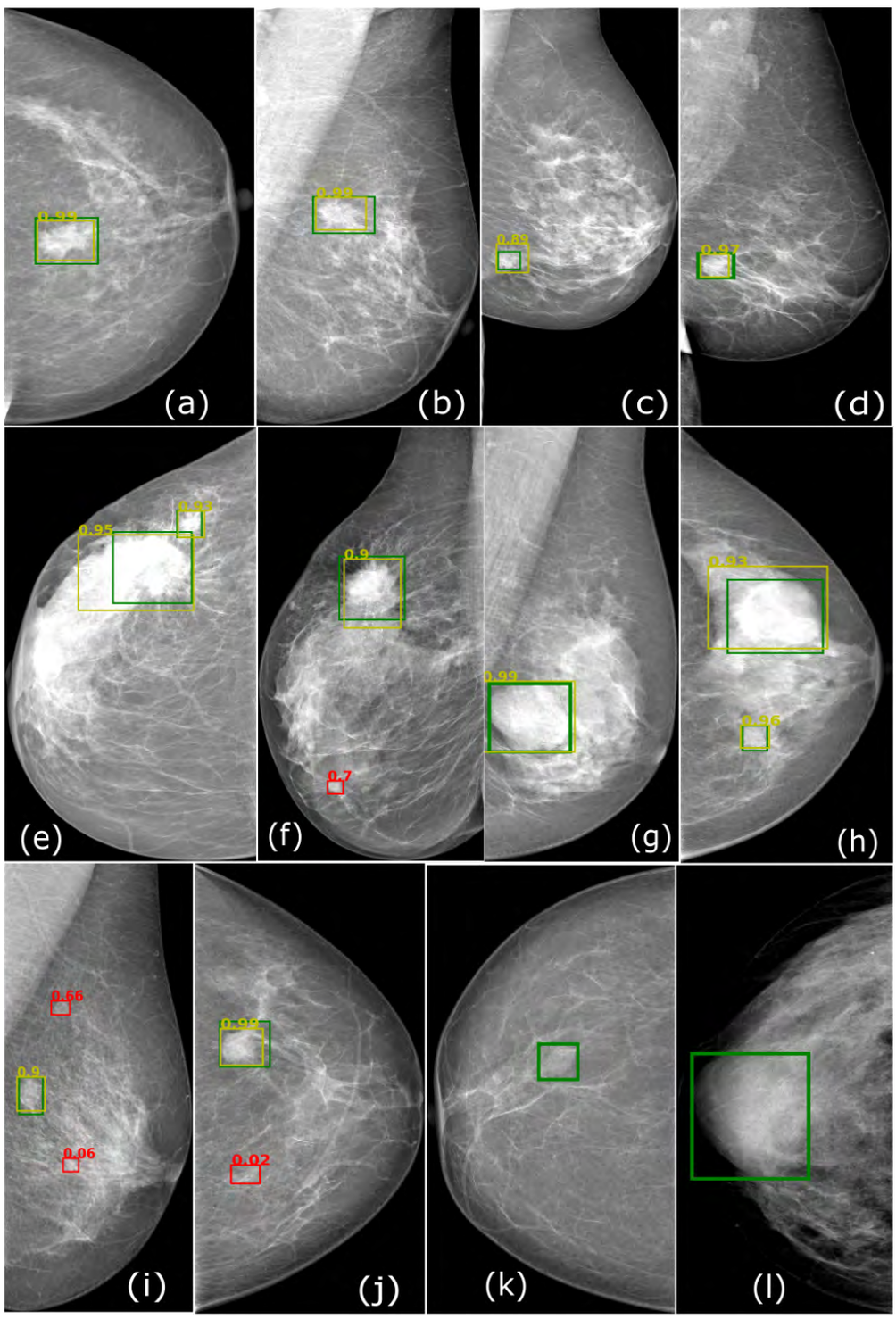


Figure 5.15: Mass detection results on INbreast dataset, (a-d,g) shows detections with high confidence score, (e,h) show multiple detections in the same mammogram, (f,i,j) shows detections with FPs, and (k,l) shows undetected masses (green: GT box, yellow and red: detection box). The numbers shown in images corresponds to the confidence of being mass.

Table 5.6: Comparison between proposed framework and results published in the literature using INbreast dataset, where μ and σ refer to the mean and standard deviation for five fold cross validation.

Methods	**TPR ($\mu \pm \sigma$) at FPI**	**# Images (INbreast)**
Kozegar et al. [90]	0.87 at 3.67	107/410
Akselrod et al. [95]	0.93 at 0.56	100/410
Dhungel et al. [31]	0.90 ± 0.02 at 1.3	410
Ribli et al. [74]	0.90 at 0.3	Malignant only
Jung et al. [98]	0.94 ± 0.05 at 1.3	410
Agarwal et al. [171]	0.98 ± 0.02 at 1.67	410
Proposed framework	0.92 ± 0.08 at 0.3 0.99 ± 0.03 **at 1.17**	Malignant
	0.71 ± 0.18 at 0.32 0.85 ± 0.08 **at 1.0**	Benign
	0.87 ± 0.05 at 0.30 0.95 ± 0.03 at 1.14	All

performed. A preliminary experiment was performed to detect lesions in the original mammograms in the OPT-G dataset (without normalization) to obtain AUROC= 0.76, which is substantially lower compared to that obtained for the normalised images (AUROC= 0.87). This can be justified as the Faster R-CNN model was pre-trained on the OPT-H images, which are similar in contrast to the normalised images in the OPT-G dataset. This contrast enhancement benefits the fine-tuning process, and as expected the system was able to detect approximately 83% of masses in the OPT-G dataset with only 0.43 FPIs, and more than 90% of masses were detected with 1.7 FPIs.

A comparison was established between the mass detection model proposed in this chapter, and the other state-of-the-art methods using the INbreast dataset as shown in Table 5.6. The mass detection model of Akselrod-Ballin et al. [95] used a modified version of the Faster R-CNN model and trained using in-house dataset containing 750 mammograms with masses. The authors obtained a TPR of 0.93 at 0.56 FPI on a subset of INbreast dataset. The model proposed by Dhungel et al. [31] consisted of a cascade of deep learning methods aimed to reduce the false positive detections and subsequently improve the precision of bounding box predictions, obtaining TPR of 0.95 ± 0.02 at 5 FPI.

Ribli et al. [74] used the Faster R-CNN model with VGG16 and used the

mammograms consisting not only the masses but also the calcifications for training and testing the model. They evaluated the performance only on the malignant lesions to obtain a TPR of 0.90 at 0.3 FPI. Jung et al. [98] used an in-house dataset (222 FFDMs) of malignant lesions along with INbreast dataset for training the RetinaNet model, obtaining a TPR of 0.94 ± 0.05 at 1.3 FPI. In the work presented in chapter 4, and also published in [171], the detection model used the mass probability map generated on each mammogram using the predictions on smaller regions (patches). The results showed superior performance with a TPR of 0.98 ± 0.02 at 1.67 FPI.

The mass detection method proposed in this chapter showed higher performance on INbreast dataset, compared to other methods in the literature detecting approximately 99% of the malignant masses with only 1.17 FPIs. These results indicate that superior performances are obtained by performing mass detection in mammograms using the recently developed deep learning based, open source object detection systems. In terms of computational efficiency, the method proposed in this chapter takes $\approx 3-5$ seconds compared to $\approx 30-50$ seconds for the traditional CNN based method in chapter 4. This shows the potential of proposed method in the development of Computer Aided Detection (CAD) systems, which could help radiologists to detect more cancers.

Chapter 6

Conclusions and Future Works

This chapter summarizes the results obtained, the main contributions and also discusses the possible future works.

6.1 Summary

The research in chapter 3 was focussed on the problem of lesion segmentation in ABUS volumes. A lesion segmentation framework was proposed for breast cancer screening in Automated Breast Ultrasound (ABUS) volumes, with the aim of supporting radiologists in breast cancer diagnosis. It was shown that the results obtained using the presented framework were less sensitive to the selection of initial seed point by the radiologist, showing the robustness of the methodology.

A volumetric analysis of the 3D segmented lesions was performed, and the results showed a high correlation between the GT and segmented volumes. For very few cases, the segmented lesion volume was greater than the GT volume, signifying that the proposed framework does not result in over segmentation of the lesions. Since, there are no publicly available ABUS dataset, after the developments in chapter 3, the further research was focussed on mammography datasets. This was done owing to the fact that still mammography is the most commonly used technique for breast cancer diagnosis in large population.

With the advent of deep learning methods, an automated methodology for detecting lesions (based on CNNs) was presented in chapter 4. In this methodology,

small regions of mammograms (patches) were used to train the CNNs. The results of training on the SFM (CBIS-DDSM) dataset demonstrated that the feature domain of the CNN can be well adapted from natural images to classify masses in mammograms.

Later, it was shown that the performance of CNN (in terms of mass detection) can be substantially enhanced by using the transfer learning between the images of similar domain (i.e. SFM→FFDM), compared to the images of different domains (natural images→mammogram). The automated mass detection framework developed in chapter 4 had shown to obtain the best results based on TPR and FPI, outperforming current state-of-the-art approaches using the same INbreast dataset.

In chapter 5, it was showed that the Faster R-CNN model pre-trained on an entirely different dataset of natural images can be adapted to efficiently detect masses in whole mammograms. Thereafter, it was shown that enhanced performances can be obtained when the Faster R-CNN model was trained on large database of mammogram and fine-tuned using the mammograms in smaller databases (OPT-G and INbreast). Compared to the other works in the literature, the proposed lesion detection framework showed improved performance in terms of higher TPR with lower FPI.

6.1.1 Contributions

The goal of this thesis was to develop efficient Computer Aided Detection (CAD) tools that can be used in clinical environment to assist radiologists in the challenging task of breast cancer detection. In this regards, the main accomplishments of this thesis can be summarized as:

1. The development of a semi-automatic framework which uses the seed point selected by the radiologists as the input and produces the segmented lesion volume as the output.

2. A volumetric temporal analysis of the breast lesions in ABUS is presented, which can be efficiently used to track the growth of lesions over time. To the best of the author's knowledge, this is the first study until November 2017 [173], presenting the volumetric temporal analysis on lesion segmentation of 3D ABUS volumes.

3. The use of CNNs in the field of breast mammography imaging has been inhibited by the limited availability of the data. In this study, it is shown that

the concept of transfer learning can be efficiently used to create automated CAD systems for analysing lesions in mammograms.

4. An efficient methodology is presented that uses only the small regions of mammograms (patches) to train the Convolutional Neural Network (CNN) and then use these predictions for detecting lesions in whole mammograms. This approach is efficient and addresses the issue of no publicly available large datasets in breast mammography imaging.

5. The successful implementation of Faster R-CNN model for detecting lesions in a large-scale dataset of breast mammograms is presented. This approach has the potential to be used within the clinical environment as it takes whole mammogram as the input and outputs the suspicious lesions within the same mammogram.

6. The presented Faster R-CNN framework has been used to detected lesions in two different small mammography datasets obtained using different scanners. This shows the potential of the detection framework to be used for analysing mammograms from different scanners, which is a peculiar requirement to be successful in different clinical environments.

6.1.2 International Research Stay

During this PhD thesis, I had the opportunity to spend 9 months (09/2018 to 06/2019) on a research stay at the Visual Computing Lab, at the Manchester Metropolitan University, Manchester, UK, under the supervision of Dr. Moi Hoon Yap. The work presented in **chapter 5** has been completed during this research period, and a journal publication is in progress.

6.2 Future Works

In this thesis, automation strategies have been proposed to develop an advanced CAD systems, which could assist radiologists in the breast cancer screening. As future work, the following extensions to the work in this thesis are proposed:

The lesion segmentation framework presented in **chapter 3** was limited by the small amount of ABUS images. As part of future work, it would be interesting to work with larger datasets in order to test the repeatability of the developed framework. Another advantage of having a large dataset would be the possible use of machine learning approaches on ABUS volumes. It would be beneficial to

develop a fully automated lesion segmentation process which could be done either by using another detection tool on the top of the framework proposed in **chapter 4** or to improve the current framework for automatic lesion detection. An application of this tool in a CAD workstation can be developed by combining the registration of the temporal volumes with this tool to perform segmentation of the lesion.

The patch classification discussed in **chapter 4** was based on the classification of the central pixel. In the future research, it would be beneficial to analyse the effect of training the CNN using the volume (i.e. no. of pixels) of lesion within each patch. Analysing the lesion detection results for Full-Field Digital Mammogram (FFDM) in **chapter 4**, it is consider that this methodology has the potential to be deployed in the hospital and can be used to assist radiologist in the breast cancer screening.

It is also consider that the methodology proposed in **chapter 5** has the potential to be adapted to detect lesions in 3D volumes such as ABUS and DBT, which are currently being adopted in the clinical practise. In this regard, an extensively large database composed of slices of 3D volumes would be required for training and testing of the CNN. This would require higher computing resources and may require using multiple GPUs for the purpose of training the CNN.

Bibliography

[1] B. Stewart and C. Wild, *World Cancer report*. World Health Organisation, 2014.

[2] F. Jacques, E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. Coebergh, H. Comber, D. Forman, and F. Bray, "Cancer incidence and mortality patterns in europe: Estimates for 40 countries in 2012", *European journal of cancer*, vol. 49, no. 6, pp. 1374–1403, 2013. DOI: 10.1016/j.ejca.2012.12.027.

[3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries", *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018. DOI: 10.3322/caac.21492.

[4] *Eurostat health statistics*, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cancer_statistics_-_specific_cancers, Accessed: 05/30/2019, 2019.

[5] J. Ferrando, L. Palència, M. Gotsens, V. Puig-Barrachina, M. Marí-Dell'Olmo, M. Rodríguez-Sanz, X. Bartoll, and C. Borrell, "Trends in cancer mortality in Spain: the influence of the financial crisis", *Gaceta Sanitaria*, vol. 33, no. 3, pp. 229–234, 2019. DOI: 10.1016/j.gaceta.2017.11.008.

[6] *European commission initiative on breast cancer*, https://ecibc.jrc.ec.europa.eu/recommendations/list/Professional, [Online; accessed 1-July-2019].

[7] *Anatomy of the breast*, https://www.mskcc.org/cancer-care/types/breast/anatomy-breast, [Online; accessed 1-July-2019].

[8] N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. A. Jong, G. Hislop, A. Chiarelli, S. Minkin, *et al.*, "Mammographic density and the risk and detection of breast cancer", *New England Journal of Medicine*, vol. 356, no. 3, pp. 227–236, 2007. DOI: 10.1056/NEJMoa062790.

[9] M. T. Mandelson, N. Oestreicher, P. L. Porter, D. White, C. A. Finder, S. H. Taplin, and E. White, "Breast density as a predictor of mammographic detection: Comparison of interval-and screen-detected cancers", *Journal of the National Cancer Institute*, vol. 92, no. 13, pp. 1081–1087, 2000. DOI: 10.1093/jnci/92.13.1081.

[10] E. D. Pisano, R. E. Hendrick, M. J. Yaffe, J. K. Baum, S. Acharyya, J. B. Cormack, L. A. Hanna, E. F. Conant, L. L. Fajardo, L. W. Bassett, *et al.*, "Diagnostic accuracy of digital versus film mammography: Exploratory analysis of selected population subgroups in dmist 1", *Radiology*, vol. 246, no. 2, pp. 376–383, 2008. DOI: 10.1148/radiol.2461070200.

[11] E. M. Fallenberg and M. Fuchsjäger, *Screening & Beyond Medical Imaging in the Detection, Diagnosis and Management of Breast Diseases*. European Society of Radiology, 2016.

[12] A. Oliver, "Automatic mass segmentation in mammographic images", PhD thesis, Universitat de Girona, 2007.

[13] *DBT*, https://radcomminc.learnupon.com/store/359919-digital-breast-tomosynthesis, [Online; accessed 22-June-2019].

[14] N. Houssami and P. Skaane, "Overview of the evidence on digital breast tomosynthesis in breast cancer detection", *The Breast*, vol. 22, no. 2, pp. 101–108, 2013. DOI: 10.1016/j.breast.2013.01.017.

[15] F. J. Gilbert, L. Tucker, and K. C. Young, "Digital breast tomosynthesis (dbt): A review of the evidence for use as a screening tool", *Clinical Radiology*, vol. 71, no. 2, pp. 141–150, 2016. DOI: 10.1016/j.crad.2015.11.008.

[16] H.-B. Pan, K.-F. Wong, A. Yao, G.-C. Hsu, C.-P. Chou, H.-L. Liang, J.-S. Huang, H.-J. Li, S.-C. Wang, and T.-L. Yang, "Breast cancer screening with digital breast tomosynthesis - 4 year experience and comparison with national data", *Journal of the Chinese Medical Association*, vol. 81, no. 1, pp. 70–80, 2018. DOI: 10.1016/j.jcma.2017.05.013.

[17] A. Hadjipanteli, P. Elangovan, A. Mackenzie, P. T. Looney, K. Wells, D. R. Dance, and K. C. Young, "The effect of system geometry and dose on the threshold detectable calcification diameter in 2d-mammography and digital breast tomosynthesis", *Physics in Medicine and Biology*, vol. 62, no. 3, pp. 858–877, Jan. 2017. DOI: 10.1088/1361-6560/aa4f6e.

[18] R. Marti, *Medical Image Analysis, VIBOT MSc Lecture Notes*, 2015.

[19] G. Pons Rodríguez, "Computer-aided lesion detection and segmentation on breast ultrasound", PhD thesis, Universitat de Girona, 2014.

[20] Y. Yan, X. Sun, and B. Shen, "Contrast agents in dynamic contrast-enhanced magnetic resonance imaging", *Oncotarget*, vol. 8, no. 26, p. 43 491, 2017. DOI: 10.18632/oncotarget.16482.

[21] E. Sickles, C. d'Orsi, L. Bassett, C. Appleton, W. Berg, E. Burnside, *et al.*, "ACR BI-RADS® mammography", *ACR BI-RADS® Atlas, Breast imaging reporting and data system*, vol. 5, 2013.

[22] M. D. Halling-Brown, P. T. Looney, M. N. Patel, L. M. Warren, A. Mackenzie, and K. C. Young, "The oncology medical image database (omi-db)", vol. 9039, 2014. DOI: 10.1117/12.2041674.

[23] *Radiology assistant*, http://www.radiologyassistant.nl/en/p53b4082c92130/bi-rads-for-mammography-and-ultrasound-2013.html, [Online; accessed 2-July-2019].

[24] R. Warren and W. Duffy, "Comparison of single reading with double reading of mammograms, and change in effectiveness with experience", *The British journal of radiology*, vol. 68, no. 813, pp. 958–962, 1995. DOI: 10.1259/0007-1285-68-813-958.

[25] J. Dinnes, S. Moss, J. Melia, R. Blanks, F. Song, and J. Kleijnen, "Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: Findings of a systematic review", *The Breast*, vol. 10, no. 6, pp. 455–463, 2001. DOI: 10.1054/brst.2001.0350.

[26] E. F. Conant, D. Li, S. Gavenonis, P. R. Bakic, A.-K. Carton, C. Zhang, A. D. Maidment, and D. Kontos, "A comparative study of the inter-reader variability of breast percent density estimation in digital mammography: Potential effect of reader's training and clinical experience", in *International Workshop on Digital Mammography*, Springer, 2010, pp. 114–120.

[27] A. Redondo, M. Comas, F. Macia, F. Ferrer, C. Murta-Nascimento, M. Maristany, E. Molins, M. Sala, and X. Castells, "Inter-and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms", *The British journal of radiology*, vol. 85, no. 1019, pp. 1465–1470, 2012. DOI: 10.1259/bjr/21256379.

[28] F. J. Gilbert, S. M. Astley, M. G. Gillan, O. F. Agbaje, M. G. Wallis, J. James, C. R. Boggis, and S. W. Duffy, "Single reading with computer-aided detection for screening mammography", *New England Journal of Medicine*, vol. 359, no. 16, pp. 1675–1684, 2008. DOI: 10.1056/NEJMoa0803545.

[29] S. S. Gómez, M. T. Tabanera, A. V. Bolivar, M. S. Miranda, A. B. Mazo, M. R. Diaz, P. M. Miravete, E. L. Asturiano, P. M. Cacho, and T. D. Macias, "Impact of a CAD system in a screen-film mammography screening program: A prospective study", *European journal of radiology*, vol. 80, no. 3, e317–e321, 2011. DOI: 10.1016/j.ejrad.2010.08.031.

[30] W. Jorritsma, F. Cnossen, and P. van Ooijen, "Improving the radiologist–CAD interaction: Designing for appropriate trust", *Clinical radiology*, vol. 70, no. 2, pp. 115–122, 2015. DOI: 10.1016/j.crad.2014.09.017.

[31] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention", *Medical image analysis*, vol. 37, pp. 114–128, 2017. DOI: 10.1016/j.media.2017.01.009.

[32] M. A. Al-antari, M. A. Al-masni, M.-T. Choi, S.-M. Han, and T.-S. Kim, "A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification", *International journal of medical informatics*, vol. 117, pp. 44–54, 2018. DOI: 10.1016/j.ijmedinf.2018.06.003.

[33] *SMARTER*, http://atc.udg.edu/~marly/smarter.html.

[34] P. Taylor, J. Champness, R. Given-Wilson, K. Johnston, and H. Potts, "Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography", 2005.

[35] J. J. Fenton, L. Abraham, S. H. Taplin, B. M. Geller, P. A. Carney, C. D'orsi, J. G. Elmore, W. E. Barlow, and B. C. S. Consortium, "Effectiveness of computer-aided detection in community mammography practice", *Journal*

of the National Cancer institute, vol. 103, no. 15, pp. 1152–1161, 2011. DOI: 10.1093/jnci/djr206.

[36] V. M. Rao, D. C. Levin, L. Parker, B. Cavanaugh, A. J. Frangos, and J. H. Sunshine, "How widely is computer-aided detection used in screening and diagnostic mammography?", *Journal of the American College of Radiology*, vol. 7, no. 10, pp. 802–805, 2010. DOI: 10.1016/j.jacr.2010.05.019.

[37] L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography", *Radiology*, vol. 215, no. 2, pp. 554–562, 2000. DOI: 10.1148/radiology.215.2.r00ma15554.

[38] A. Malich, C. Marx, M. Facius, T. Boehm, M. Fleck, and W. A. Kaiser, "Tumour detection rate of a new commercially available computer-aided detection system", *European radiology*, vol. 11, no. 12, pp. 2454–2459, 2001. DOI: 10.1007/s003300101079.

[39] J. A. Baker, J. Y. Lo, D. M. Delong, and C. E. Floyd, "Computer-aided detection in screening mammography: Variability in cues", *Radiology*, vol. 233, no. 2, pp. 411–417, 2004. DOI: 10.1148/radiol.2332031200.

[40] D. R. Chen, R. F. Chang, W. J. Wu, W. K. Moon, and W. L. Wu, "3-D breast ultrasound segmentation using active contour model", *Ultrasound in medicine & biology*, vol. 29, no. 7, pp. 1017–1026, 2003. DOI: 10.1016/S0301-5629(03)00059-0.

[41] T. Tan, B. Platel, H. Huisman, C. I. Sanchez, R. Mus, and N. Karssemeijer, "Computer-aided lesion diagnosis in automated 3-d breast ultrasound using coronal spiculation", *IEEE Transactions on Medical Imaging*, vol. 31, no. 5, pp. 1034–1042, 2012. DOI: 10.1109/TMI.2012.2184549.

[42] W. K. Moon, C. M. Lo, J. M. Chang, C. S. Huang, J. H. Chen, and R. F. Chang, "Computer-aided classification of breast masses using speckle features of automated breast ultrasound images", *Medical Physics*, vol. 39, no. 10, pp. 6465–6473, 2012. DOI: 10.1118/1.4754801.

[43] *ASSURE*, www.assure-project.eu.

[44] T. Tan, A. Gubern-Mérida, C. Borelli, R. Manniesing, J. van Zelst, L. Wang, W. Zhang, B. Platel, R. M. Mann, and N. Karssemeijer, "Segmentation of malignant lesions in 3d breast ultrasound using a depth-dependent model", *Medical Physics*, vol. 43, no. 7, pp. 4074–4084, 2016. DOI: 10.1118/1.4953206.

[45] D. Lowe, "Object recognition from local scale-invariant features", in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, Sep. 1999, pp. 1150–1157. DOI: 10.1109/ICCV.1999.790410.

[46] *ITK-SNAP*, www.itksnap.org.

[47] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, "The digital database for screening mammography", *Digital mammography*, pp. 431–434, 2000.

[48] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research", *Scientific data*, vol. 4, p. 170 177, 2017. DOI: 10.1038/sdata.2017.177.

[49] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a full-field digital mammographic database", *Academic Radiology*, vol. 19, no. 2, pp. 236–248, 2012. DOI: 10.1016/j.acra.2011.09.014.

[50] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, *et al.*, "The mammographic image analysis society digital mammogram database", in *International Congress Series*, Excerta Medica, 1994.

[51] M. G. Lopez, N. G. Posada, D. C. Moura, R. R. Pollán, J. M. F. Valiente, C. S. Ortega, M. Solar, G. Diaz-Herrero, I. Ramos, J. Loureiro, *et al.*, "BCDR: A breast cancer digital repository", in *15th International conference on experimental mechanics*, 2012.

[52] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, *et al.*, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository", *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013. DOI: 10.1007/s10278-013-9622-7.

[53] M. Samulski, R. Hupse, C. Boetes, R. D. Mus, G. J. den Heeten, and N. Karssemeijer, "Using computer-aided detection in mammography as a decision support", *European radiology*, vol. 20, no. 10, pp. 2323–2330, 2010. DOI: 10.1007/s00330-010-1821-8.

[54] E. D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J. K. Baum, S. Acharyya, E. F. Conant, L. L. Fajardo, L. Bassett, C. DÓrsi, *et al.*, "Diagnostic performance of digital versus film mammography for breast-cancer screening", *New England Journal of Medicine*, vol. 353, no. 17, pp. 1773–1783, 2005. DOI: 10.1056/NEJMoa052911.

[55] G. Pons, J. Martí, R. Martí, S. Ganau, J. C. Vilanova, and J. A. Noble, "Evaluating lesion segmentation on breast sonography as related to lesion type", *Journal of Ultrasound in Medicine*, vol. 32, no. 9, pp. 1659–1670, 2013. DOI: 10.7863/ultra.32.9.1659.

[56] G. Pons, J. Martí, R. Martí, and J. A. Noble, "Simultaneous lesion segmentation and bias correction in breast ultrasound images", in *Iberian Conference on Pattern Recognition and Image Analysis*, Springer, 2011, pp. 692–699.

[57] J. Massich, F. Meriaudeau, E. Pérez, R. Martí, A. Oliver, and J. Martí, "Lesion segmentation in breast sonography", vol. 6136, 2010, pp. 39–45.

[58] A. Nugroho, H. A. Nugroho, and L. Choridah, "Active contour bilateral filter for breast lesions segmentation on ultrasound images", in *2015 International Conference on Science in Information Technology (ICSITech)*, IEEE, 2015, pp. 36–40.

[59] R. F. Chang, W. J. Wu, W. K. Moon, and D. R. Chen, "Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors", *Breast Cancer Research and Treatment*, vol. 89, no. 2, pp. 179–185, 2005. DOI: 10.1007/s10549-004-2043-z.

[60] K. M. Kelly, J. Dean, W. S. Comulada, and S.-J. Lee, "Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts", *European Radiology*, vol. 20, no. 3, pp. 734–742, 2010. DOI: 10.1007/s00330-009-1588-y.

[61] T. Tan, B. Platel, H. Huisman, C. I. Sánchez, R. Mus, and N. Karssemeijer, "Computer-aided lesion diagnosis in automated 3-D breast ultrasound using coronal spiculation", *IEEE Transactions on Medical Imaging*, vol. 31, no. 5, pp. 1034–1042, 2012. DOI: 10.1109/TMI.2012.2184549.

[62] T. Tan, A. Gubern-Mérida, C. Borelli, R. Manniesing, J. van Zelst, L. Wang, W. Zhang, B. Platel, R. M. Mann, and N. Karssemeijer, "Segmentation of malignant lesions in 3D breast ultrasound using a depth-dependent model", *Medical Physics*, vol. 43, no. 7, pp. 4074–4084, 2016. DOI: 10.1118/1.4953206.

[63] O. Diaz, R. Agarwal, A. Gubern-Mérida, J. Zelst Van, D. Yago, and R. Martí, "Automated volumetric lesion quantification in automated 3D breast ultrasound: Comparison of 5 breast lesion segmentation algorithms", in, 1, vol. 7, 2016, pp. 162–465.

[64] J. C. van Zelst, T. Tan, P. Clauser, A. Domingo, M. D. Dorrius, D. Drieling, M. Golatta, F. Gras, M. de Jong, R. Pijnappel, *et al.*, "Dedicated computer-aided detection software for automated 3D breast ultrasound; an efficient tool for the radiologist in supplemental screening of women with dense breasts", *European radiology*, vol. 28, no. 7, pp. 2996–3006, 2018. DOI: 10.1007/s00330-017-5280-3.

[65] J. Canny, "A computational approach to edge detection", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 6, pp. 679–698, 1986. DOI: 10.1109/TPAMI.1986.4767851.

[66] N. Otsu, "A threshold selection method from gray-level histograms", *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975. DOI: 10.1109/TSMC.1979.4310076.

[67] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Pearson Education Inc. Upper Saddle River, New Jersey, 2008.

[68] C. Lo, R. Chen, Y. Chang, Y. Yang, M. Hung, C. Huang, and R. Chang, "Multi-dimensional tumor detection in automated whole breast ultrasound using topographic watershed", *IEEE transactions on medical imaging*, vol. 33, no. 7, pp. 1503–1511, 2014. DOI: 10.1109/TMI.2014.2315206.

[69] R. Agarwal, "A Semi Automatic Tool for Breast Lesion Segmentation in 3D-ABUS Volumes", Master's thesis, Universitat de Girona, Girona, 2016.

[70] N. Dhungel, G. Carneiro, and A. P. Bradley, "Automated mass detection in mammograms using cascaded deep learning and random forests", in *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, 2015, pp. 1–8.

[71] G. Carneiro, J. Nascimento, and A. P. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 652–660.

[72] W. Lotter, G. Sorensen, and D. Cox, "A multi-scale CNN and curriculum learning strategy for mammogram classification", ser. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2017, pp. 169–177.

[73] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C. I. Sánchez, R. Mann, A. den Heeten, and N. Karssemeijer, "Large scale deep learning for computer aided detection of mammographic lesions", *Medical image analysis*, vol. 35, pp. 303–312, 2017. DOI: 10.1016/j.media.2016.07.007.

[74] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with deep learning", *Scientific reports*, vol. 8, no. 1, p. 4165, 2018. DOI: 10.1038/s41598-018-22437-z.

[75] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. DOI: 10.1038/nature14539.

[76] R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection", *Radiology*, vol. 219, no. 1, pp. 192–202, 2001. DOI: 10.1148/radiology.219.1.r01ap16192.

[77] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", in *Proceedings of the conference on computer vision and pattern recognition*, IEEE, 2014, pp. 580–587.

[78] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection", in *Advances in neural information processing systems*, NIPS, 2013, pp. 2553–2561.

[79] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017. DOI: 10.1109/tpami.2016.2577031.

[80] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in *Advances in neural information processing systems*, NIPS, 2012, pp. 1097–1105.

[81] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", in *Proceedings of the conference on computer vision and pattern recognition*, IEEE, 2015, pp. 1–9.

[82] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1915–1929, 2013. DOI: 10.1109/TPAMI.2012.231.

[83] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation", in *Advances in neural information processing systems*, NIPS, 2014, pp. 1799–1807.

[84] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch", *Journal of Machine Learning Research*, vol. 12, pp. 2493–2537, 2011.

[85] A. Bordes, S. Chopra, and J. Weston, "Question answering with subgraph embeddings", in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, EMNLP, 2014, pp. 615–620.

[86] T. Mikolov, A. Deoras, D. Povey, L. Burget, and J. Černocký, "Strategies for training large scale neural network language models", in *Automatic Speech Recognition and Understanding*, IEEE, 2011, pp. 196–201.

[87] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, and T. N. Sainath, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups", *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012. DOI: 10.1109/MSP.2012.2205597.

[88] M. L. Giger, N. Karssemeijer, and J. A. Schnabel, "Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer", *Annual review of biomedical engineering*, vol. 15, pp. 327–357, 2013. DOI: 10.1146/annurev-bioeng-071812-152416.

[89] A. Oliver, J. Freixenet, J. Marti, E. Perez, J. Pont, E. R. Denton, and R. Zwiggelaar, "A review of automatic mass detection and segmentation in mammographic images", *Medical image analysis*, vol. 14, no. 2, pp. 87–110, 2010. DOI: 10.1016/j.media.2009.12.005.

[90] E. Kozegar, M. Soryani, B. Minaei, and I. Domingues, "Assessment of a novel mass detection algorithm in mammograms", *Journal of cancer research and therapeutics*, vol. 9, no. 4, p. 592, 2013. DOI: 10.4103/0973-1482.126453.

[91] A. Becker, M. Marcon, S. Ghafoor, M. C. Wurnig, T. Frauenfelder, and A. Boss, "Deep learning in mammography: Diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer", eng, *Investigative radiology*, vol. 52, no. 7, pp. 434–440, 2017. DOI: 10.1097/RLI.0000000000000358.

[92] S. Vesal, S. M. Patil, N. Ravikumar, and A. K. Maier, "A multi-task framework for skin lesion detection and segmentation", in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, Springer, 2018, pp. 285–293.

[93] M. Goyal, N. Reeves, S. Rajbhandari, and M. H. Yap, "Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices", *IEEE journal of biomedical and health informatics*, 2018. DOI: 10.1109/JBHI.2018.2868656.

[94] A. Akselrod-Ballin, L. Karlinsky, S. Alpert, S. Hashoul, R. Ben-Ari, and E. Barkan, "A CNN based method for automatic mass detection and classification in mammograms", *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 7, no. 3, pp. 242–249, 2019. DOI: 10.1080/21681163.2017.1350206.

[95] A. Akselrod-Ballin, L. Karlinsky, A. Hazan, R. Bakalo, A. B. Horesh, Y. Shoshan, and E. Barkan, "Deep learning for automatic detection of abnormal findings in breast mammography", in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2017, pp. 321–329.

[96] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets", *CoRR*, vol. abs/1405.3531, 2014. arXiv: 1405.3531.

[97] S. Zagoruyko and N. Komodakis, "Wide residual networks", *CoRR*, vol. abs/1605.07146, 2016. arXiv: 1605.07146.

[98] H. Jung, B. Kim, I. Lee, M. Yoo, J. Lee, S. Ham, O. Woo, and J. Kang, "Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network", *PloS one*, vol. 13, no. 9, e0203355, 2018. DOI: 10.1371/journal.pone.0203355.

[99] S. Morrell, Z. Wojna, C. S. Khoo, S. Ourselin, and J. E. Iglesias, "Large-scale mammography CAD with deformable conv-nets", in *Image Analysis for Moving Organ, Breast, and Thoracic Images*, Springer, 2018, pp. 64–72.

[100] M. A. Al-masni, M. A. Al-antari, J.-M. Park, G. Gi, T.-Y. Kim, P. Rivera, E. Valarezo, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLOs-based CAD system", *Computer methods and programs in biomedicine*, vol. 157, pp. 85–94, 2018. DOI: 10.1016/j.cmpb.2018.01.017.

[101] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection", in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[102] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[103] W. A. Berg, J. D. Blume, J. B. Cormack, and E. B. Mendelson, "Operator dependence of physician-performed whole-breast us: Lesion detection and characterization 1", *Radiology*, vol. 241, no. 2, pp. 355–365, 2006. DOI: 10.1148/radiol.2412051710.

[104] H. J. Shin, H. H. Kim, and J. H. Cha, "Current status of automated breast ultrasonography", *Ultrasonography*, vol. 34, no. 3, p. 165, 2015. DOI: 10.14366/usg.15002.

[105] P. Therasse, S. G. Arbuck, E. A. Eisenhauer, J. Wanders, R. S. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A. T. van Oosterom, M. C. Christian, *et al.*, "New guidelines to evaluate the response to treatment in solid tumors", *Journal of the National Cancer Institute*, vol. 92, no. 3, pp. 205–216, 2000. DOI: 10.1093/jnci/92.3.205.

[106] T. Heinonen, P. Dastidar, H. Eskola, H. Frey, P. Ryymin, and E. Laasonen, "Applicability of semi-automatic segmentation for volumetric analysis of brain lesions", *Journal of medical engineering & technology*, vol. 22, no. 4, pp. 173–178, 1998. DOI: 10.3109/03091909809032536.

[107] S. Valverde, A. Oliver, and X. Lladó, "A white matter lesion-filling approach to improve brain tissue volume measurements", *NeuroImage: Clinical*, vol. 6, pp. 86–92, 2014. DOI: 10.1016/j.nicl.2014.08.016.

[108] E. Roura, N. Sarbu, A. Oliver, S. Valverde, S. González-Villá, R. Cervera, N. Bargalló, and X. Lladó, "Automated detection of lupus white matter lesions in MRI", *Frontiers in Neuroinformatics*, vol. 10, p. 33, 2016. DOI: 10.3389/fninf.2016.00033.

[109] P. D. Mozley, C. Bendtsen, B. Zhao, L. H. Schwartz, M. Thorn, Y. Rong, L. Zhang, A. Perrone, R. Korn, and A. J. Buckler, "Measurement of tumor volumes improves recist-based response assessments in advanced lung cancer", *Translational oncology*, vol. 5, no. 1, pp. 19–25, 2012. DOI: 10.1593/tlo.11232.

[110] M. Nishino, S. E. Dahlberg, S. Cardarella, D. M. Jackman, M. S. Rabin, N. H. Ramaiya, H. Hatabu, P. A. Jänne, and B. E. Johnson, "Volumetric tumor growth in advanced non-small cell lung cancer patients with EGFR mutations during EGFR-tyrosine kinase inhibitor therapy", *Cancer*, vol. 119, no. 21, pp. 3761–3768, 2013. DOI: 10.1002/cncr.28290.

[111] E. J. Kim, S. H. Kim, B. J. Kang, and Y. J. Kim, "Interobserver agreement on the interpretation of automated whole breast ultrasonography", *Ultrasonography*, vol. 33, no. 4, p. 252, 2014.

[112] N. Masciadri and C. Ferranti, "Benign breast lesions: Ultrasound", *Journal of ultrasound*, vol. 14, no. 2, pp. 55–65, 2011.

[113] H. J. Johnson, M. M. McCormick, and L. Ibanez, "The ITK software guide book 2: Design and functionality fourth edition updated for ITK version 4.7", *Insight Journal*, 2015.

[114] M. Jiang and D. Crookes, "High-performance 3d median filter architecture for medical image despeckling", *Electronics Letters*, vol. 42, no. 24, pp. 1379–1380, Nov. 2006. DOI: 10.1049/el:20062357.

[115] I. T. Young and L. J. Van Vliet, "Recursive implementation of the gaussian filter", *Signal processing*, vol. 44, no. 2, pp. 139–151, 1995. DOI: 10.1016/0165-1684(95)00020-E.

[116] I. Macía, *Generalized computation of gaussian derivatives using ITK*, 2007.

[117] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 7, pp. 629–639, 1990. DOI: 10.1109/34.56205.

[118] J. Weickert, *Anisotropic diffusion in image processing*. Image Rochester NY, 1998, vol. 256, p. 170.

[119] J. Fehrenbach and J.-M. Mirebeau, "Sparse non-negative stencils for anisotropic diffusion", *Journal of Mathematical Imaging and Vision*, vol. 49, no. 1, pp. 123–147, 2013. DOI: 10.1007/s10851-013-0446-3.

[120] J. Mirebeau, J. Fehrenbach, L. Risser, and S. Tobji, "Anisotropic diffusion in ITK", *Insight Journal*, Dec. 2014.

[121] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Pearson Education, 2008.

[122] R. Beare and G. Lehmann, "The watershed transform in ITK-discussion and new developments", *Insight J*, 2006.

[123] E. Steve, *The watershed transform: Strategies for image segmentation*, http://www.mathworks.com/company/newsletters/articles/the-watershed-transform-strategies-for-image-segmentation.html, 2002.

[124] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig, "User guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability", *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006. DOI: 10.1016/j.neuroimage.2006.01.015.

[125] L. R. Dice, "Measures of the amount of ecologic association between species", *Ecology*, vol. 26, no. 3, pp. 297–302, 1945. DOI: 10.2307/1932409.

[126] W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized Overlap Measures for Evaluation and Validation in medical image analysis", *IEEE Transactions on Medical Imaging*, vol. 25, no. 11, pp. 1451–1461, 2006. DOI: 10.1109/TMI.2006.880587.

[127] N. Wang, C. Bian, Y. Wang, M. Xu, C. Qin, X. Yang, T. Wang, A. Li, D. Shen, and D. Ni, "Densely deep supervised networks with threshold loss for cancer detection in automated breast ultrasound", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, pp. 641–648.

[128] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks", in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489–4497.

[129] T.-C. Chiang, Y.-S. Huang, R.-T. Chen, C.-S. Huang, and R.-F. Chang, "Tumor detection in automated breast ultrasound using 3D CNN and prioritized candidate aggregation", *IEEE transactions on medical imaging*, vol. 38, no. 1, pp. 240–249, 2018.

[130] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation", *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1170–1181, 2016. DOI: 10.1109/TMI.2015.2482920.

[131] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification", in *International Conference on Information Processing in Medical Imaging*, Springer, 2015, pp. 588–599.

[132] A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. Van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, and B. van Ginneken, "Pulmonary nodule detection in ct images: False positive reduction using multiview convolutional networks", *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1160–1169, 2016. DOI: 10.1109/TMI.2016.2536809.

[133] A. Gupta, M. Ayhan, and A. Maida, "Natural image bases to represent neuroimaging data", in *International conference on machine learning*, 2013, pp. 987–994.

[134] T. Brosch, R. Tam, A. D. N. Initiative, *et al.*, "Manifold learning of brain MRIs by deep learning", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2013, pp. 633–640.

[135] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?", in *Advances in neural information processing systems*, NIPS, 2014, pp. 3320–3328.

[136] H. Azizpour, A. S. Razavian, J. Sullivan, A. Maki, and S. Carlsson, "From generic to specific deep representations for visual recognition", in *CVPRW DeepVision Workshop, June 11, 2015, Boston, MA, USA*, IEEE, 2015.

[137] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?", *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016. DOI: 10.1109/TMI.2016.2535302.

[138] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", in *International Conference on Learning Representations*, ICLR, 2015.

[139] S. T. Kim, H. Lee, H. G. Kim, and Y. M. Ro, "ICADx: Interpretable computer aided diagnosis of breast masses", in *Medical Imaging 2018: Computer-Aided Diagnosis*, International Society for Optics and Photonics, vol. 10575, 2018, p. 1 057 522.

[140] S. J. S. Gardezi, M. Awais, I. Faye, and F. Meriaudeau, "Mammogram classification using deep learning features", in *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, IEEE, 2017, pp. 485–488.

[141] Y. Zheng, C. Yang, and A. Merkulov, "Breast cancer screening using convolutional neural network and follow-up digital mammography", in *Computational Imaging III*, International Society for Optics and Photonics, vol. 10669, 2018, p. 1 066 905.

[142] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep convolutional neural networks for breast cancer screening", *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 19–30, 2018. DOI: 10.1016/j.cmpb.2018.01.011.

[143] S. Agrawal, R. Rangnekar, D. Gala, S. Paul, and D. Kalbande, "Detection of breast cancer from mammograms using a hybrid approach of deep learning and linear classification", in *2018 International Conference on Smart City and Emerging Technology (ICSCET)*, IEEE, 2018, pp. 1–6.

[144] X. Gu, Z. Shi, and J. Ma, "Multi-view learning for mammogram analysis: Auto-diagnosis models for breast cancer", in *2018 IEEE International Conference on Smart Internet of Things (SmartIoT)*, IEEE, 2018, pp. 149–153.

[145] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", in *Proceedings of the conference on computer vision and pattern recognition*, IEEE, 2016, pp. 770–778.

[146] L. Shen, "End-to-end training for whole image breast cancer diagnosis using an all convolutional design", *arXiv preprint arXiv:1708.09427*, 2017.

[147] F. Gao, T. Wu, J. Li, B. Zheng, L. Ruan, D. Shang, and B. Patel, "SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis", *Computerized Medical Imaging and Graphics*, vol. 70, pp. 53–62, 2018. DOI: 10.1016/j.compmedimag.2018.09.004.

[148] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision", in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, 2016, pp. 2818–2826.

[149] D. Soriano, C. Aguilar, I. Ramirez-Morales, E. Tusa, W. Rivas, and M. Pinta, "Mammogram classification schemes by using convolutional neural networks", in *International Conference on Technology Trends*, Springer, 2017, pp. 71–85.

[150] E. Moya, E. Campoverde, E. Tusa, I. Ramirez-Morales, W. Rivas, and B. Mazon, "Multi-category classification of mammograms by using convolutional neural networks", in *2017 International Conference on Information Systems and Computer Science*, IEEE, 2017, pp. 133–140.

[151] Y. Mednikov, S. Nehemia, B. Zheng, O. Benzaquen, and D. Lederman, "Transfer representation learning using Inception-V3 for the detection of masses in mammography", in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2018, pp. 2587–2590.

[152] D. P. Kingma and J. L. Ba, "ADAM: A method for stocastic optimization", in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, ACM, 2014.

[153] te Brake GM, N. Karssemeijer, and J. H. Hendriks, "An automatic method to discriminate malignant masses from normal tissue in digital mammograms", *Physics in Medicine & Biology*, vol. 45, no. 10, p. 2843, 2000. DOI: 10.1088/0031-9155/45/10/308.

[154] R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, and M. Roffilli, "A novel featureless approach to mass detection in digital mammograms based on support vector machines", *Physics in Medicine & Biology*, vol. 49, no. 6, p. 961, 2004. DOI: 10.1088/0031-9155/49/6/007.

[155] R. Agarwal, O. Diaz, X. Lladó, and R. Martí, "Mass detection in mammograms using pre-trained deep learning models", vol. 10718, Proc.SPIE, 2018, 107181F. DOI: 10.1117/12.2317681.

[156] C. Gu, J. Lim, P. Arbelaez, and J. Malik, "Recognition using regions", in *2009 IEEE Conference on Computer Vision and Pattern Recognition*.

[157] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition", *International journal of computer vision*, vol. 104, no. 2, pp. 154–171, 2013. DOI: 10.1007/s11263-013-0620-5.

[158] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context", *Lecture Notes in Computer Science*, pp. 740–755, 2014. DOI: 10.1007/978-3-319-10602-1_48.

[159] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[160] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation", *International journal of computer vision*, vol. 59, no. 2, pp. 167–181, 2004. DOI: 10.1023/B:VISI.0000022288.19776.77.

[161] R. Girshick, "Fast R-CNN", in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[162] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, "Tensorflow: A system for large-scale machine learning", in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.

[163] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and et al., "Speed/accuracy trade-offs for modern convolutional object detectors", *IEEE Conference on Computer Vision and Pattern Recognition*, Jul. 2017. DOI: 10.1109/cvpr.2017.351.

[164] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge", *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010. DOI: 10.1007/s11263-009-0275-4.

[165] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite", in *Conference on Computer Vision and Pattern Recognition*, 2012.

[166] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Malloci, T. Duerig, and V. Ferrari, *The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale*, 2018. arXiv: 1811.00982 [cs.CV].

[167] G. V. Horn, O. M. Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, and S. Belongie, "The iNaturalist Species Classification and Detection Dataset", *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018. DOI: 10.1109/cvpr.2018.00914.

[168] http://dicom.nema.org/dicom/2013/output/chtml/part16/sect_CID_4015.html, Accessed: 04/28/2019, 2019.

[169] R. Bellotti, F. De Carlo, S. Tangaro, G. Gargano, G. Maggipinto, M. Castellano, R. Massafra, D. Cascio, F. Fauci, R. Magro, *et al.*, "A completely automated CAD system for mass detection in a large mammographic database", *Medical physics*, vol. 33, no. 8, pp. 3066–3075, 2006. DOI: 10.1118/1.2214177.

[170] M. P. Sampat, A. C. Bovik, G. J. Whitman, and M. K. Markey, "A model-based framework for the detection of spiculated masses on mammography", *Medical physics*, vol. 35, no. 5, pp. 2110–2123, 2008. DOI: 10.1118/1.2890080.

[171] R. Agarwal, O. Diaz, X. Lladó, M. H. Yap, and R. Martí, "Automatic mass detection in mammograms using deep convolutional neural networks", *Journal of Medical Imaging*, vol. 6, no. 3, 2019. DOI: 10.1117/1.JMI.6.3.031409.

[172] http://dicom.nema.org/medical/dicom/current/output/chtml/part03/sect_C.11.2.html, Accessed: 05/06/2019, 2019.

[173] R. Agarwal, O. Diaz, X. Lladó, A. Gubern-Mérida, J. C. Vilanova, and R. Martí, "Lesion Segmentation in Automated 3D Breast Ultrasound: Volumetric Analysis", *Ultrasonic imaging*, vol. 40, no. 2, pp. 97–112, 2018. DOI: 10.1177/0161734617737733.