


Please cite the Published Version

Agarwal, R, Díaz, O, Yap, Moi Hoon , Lladó, X and Martí, R (2020) Deep learning for mass detection in Full Field Digital Mammograms. *Computers in Biology and Medicine*, 121. p. 103774. ISSN 0010-4825

DOI: <https://doi.org/10.1016/j.combiomed.2020.103774>

Publisher: Elsevier

Version: Published Version

Downloaded from: <https://e-space.mmu.ac.uk/625750/>

Usage rights:  [Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Additional Information: This is an Open Access article published in *Computers in Biology and Medicine*, by Elsevier.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)



Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/complbiomed

Deep learning for mass detection in Full Field Digital Mammograms

Richa Agarwal^a, Oliver Díaz^{a,b}, Moi Hoon Yap^c, Xavier Lladó^a, Robert Martí^{a,*}^a VICOROB, Department of Computer Architecture and Technology, University of Girona, Spain^b Department of Mathematics and Computer Science, University of Barcelona, Spain^c Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom

ARTICLE INFO

Keywords:

Deep learning
CNN
Mammogram
FFDM
Mass detection

ABSTRACT

In recent years, the use of Convolutional Neural Networks (CNNs) in medical imaging has shown improved performance in terms of mass detection and classification compared to current state-of-the-art methods. This paper proposes a fully automated framework to detect masses in Full-Field Digital Mammograms (FFDM). This is based on the Faster Region-based Convolutional Neural Network (Faster-RCNN) model and is applied for detecting masses in the large-scale OPTIMAM Mammography Image Database (OMI-DB), which consists of ~80,000 FFDMs mainly from Hologic and General Electric (GE) scanners. This research is the first to benchmark the performance of deep learning on OMI-DB. The proposed framework obtained a True Positive Rate (TPR) of 0.93 at 0.78 False Positive per Image (FPI) on FFDMs from the Hologic scanner. Transfer learning is then used in the Faster R-CNN model trained on Hologic images to detect masses in smaller databases containing FFDMs from the GE scanner and another public dataset INbreast (Siemens scanner). The detection framework obtained a TPR of 0.91 ± 0.06 at 1.69 FPI for images from the GE scanner and also showed higher performance compared to state-of-the-art methods on the INbreast dataset, obtaining a TPR of 0.99 ± 0.03 at 1.17 FPI for malignant and 0.85 ± 0.08 at 1.0 FPI for benign masses, showing the potential to be used as part of an advanced CAD system for breast cancer screening.

1. Introduction

Breast cancer is the most common form of cancer in the female population. It is estimated that approximately 12% of women in the USA will be diagnosed with breast cancer at some point during their lifetime [1]. Breast cancer has the highest incidence and mortality rate amongst all cancers (excluding melanoma skin cancer) [2]. In the EU, breast cancer is the leading cause of mortality amongst the female population, accounting for 15.6% in 2015 [3].

Although breast cancer incidence has increased in the past decade, the introduction of screening programmes for early detection has achieved a lowering of the mortality rate. The conventional imaging modality used for screening is X-ray mammography, being both a fast and cost-effective technique for screening a large population. In this technique, images of each breast are typically acquired using two different views: cranio-caudal (CC) imaged from top to bottom and medio-lateral oblique (MLO) from left to right. With advancements in imaging techniques, high quality full-field digital mammograms (FFDM) have replaced the traditional scanned-film mammograms. Furthermore, developments in computer technology and data science has

generated interest in exploring deep learning methods for various tasks including object detection [4,5] and image recognition [6,7].

Deep learning methods based on convolutional neural networks (CNN) have also gained importance in the field of medical image analysis and efforts have been made to develop modern computer-aided detection (CAD) systems based on these CNN algorithms [8–10]. Additionally, in mammography, some authors have also proposed the use of traditional machine learning based on handcrafted feature to classify masses [11–14]. However, the exploration of deep learning methods in the field of breast imaging has been limited, as only a small number of public datasets are available (e.g. DDSM [15], INbreast [16]).

Although researchers have used Faster R-CNN in medical imaging [17,18], there is a dearth of literature in the breast imaging field. For instance, Akselrod-Ballin et al. [19] used a modified version of a Faster R-CNN model to include information from the finer bottom levels during the classification stage. Ribli et al. [20] trained a Faster R-CNN model on the DDSM database composed of 2620 scanned-film mammograms and then evaluated the performance of the network in the INbreast dataset of malignant masses.

* Corresponding author.

E-mail addresses: agarwalrichi13@gmail.com (R. Agarwal), oliver.diaz@ub.edu (O. Díaz), m.yap@mmu.ac.uk (M.H. Yap), xavier.llado@udg.edu (X. Lladó), robert.marti@udg.edu (R. Martí).

<https://doi.org/10.1016/j.complbiomed.2020.103774>

Received 9 January 2020; Received in revised form 19 April 2020; Accepted 19 April 2020

Available online 22 April 2020

0010-4825/© 2020 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Jung et al. [21] proposed a mass detection model based on RetinaNet [22] using a new loss function, called focal loss, to address the problem of extreme class imbalance between the foreground and background. The performance of the network was evaluated on a combination of a private (GURO) and public (INbreast) dataset. Morrel et al. [23] presented a neural network based on a region-based fully convolutional network (R-FCN) [24] and deformable convolutional nets. Although the network was trained using the OPTIMAM Mammography Image Database (OMI-DB) [25], the results were only provided for the DREAMS challenge [26] competitive phase.

Recently, Al-masni et al. [9] and Al-antari et al. [10] adopted the You Only Look Once (YOLO) deep learning method [27] for the detection and classification of masses in mammograms. One of the major advantages of the YOLO algorithm is speed, as it defines the object detection as a regression problem. However, YOLO is limited in terms of accuracy and precision in the localisation of small objects [27].

In our previous work [28], an automated mass detection framework using CNN was presented. Here small regions of the mammograms (patches) were extracted using a sliding window approach and used for training different CNNs. The framework obtained results comparable to state-of-the-art on the INbreast dataset. However, the high computational cost was a limiting factor for clinical use.

In this paper, a mass detection framework based on Faster R-CNN object detection model is presented. This uses the whole FFDm (instead of patch-based strategy) for training and testing and is based on the “recognition using regions” paradigm [29]. The proposed framework is evaluated on the large mammography OMI-DB [25] dataset containing images from 4750 cases. The key contributions of this paper are summarised below:

1. The implementation of Faster R-CNN model for detecting masses in a large-scale mammography dataset of malignant masses (OMI-DB).
2. To the best of our knowledge, this paper is the first to benchmark the performance of a deep learning method on the OMI-DB dataset.
3. The application of transfer learning to detect masses in two small mammography datasets obtained using different scanners.
4. Comparison with other works in the literature, showing that the proposed mass detection framework achieves better performance in terms of higher True Positive Rate (TPR) with lower False Positives per Image (FPI).

The remainder of this paper is structured as follows: Section 2 details the datasets used; Section 3 describes the Faster R-CNN model; Section 4 describes the methodology for training and testing the Faster R-CNN model, and Section 5 details the experimental results. Discussions are presented in Section 6, and in Section 7 we provide conclusions and suggested future work.

2. Datasets

2.1. OPTIMAM mammography database (OMI-DB)

The OMI-DB [25] is an extensive mammography image database of over 145,000 cases (over 2.4 million images) comprised of unprocessed and processed FFDms from the UK’s National Health Service Breast Screening Program. It also contains expert’s determined ground truths and associated clinical data linked to the images. As part of the data sharing agreement with the Royal Surrey County Hospital (UK) in 2017, we obtained a subset of this database (4750 cases with ~80,000 processed and unprocessed FFDms). The database contains images from different manufacturers, particularly Hologic Inc, Marlborough, Massachusetts, USA (Hologic Lorad Selenia and Selenia Dimensions Mammography Systems), and General Electric (GE) Medical Systems, Chicago, Illinois, USA (Senograph DS and Senographe Essential), referred to as OMI-H and OMI-G, respectively. For each case, two views

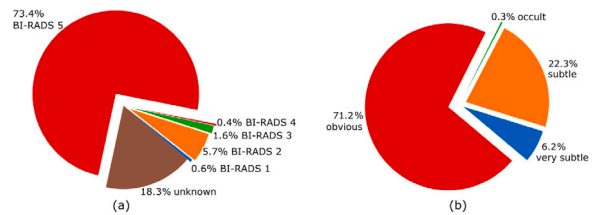


Fig. 1. Categorisation of the OMI-DB dataset using (a) BI-RADS ratings, and (b) mass conspicuity.

of each breast, i.e. medio-lateral oblique (MLO) and cranio-caudal (CC) are available, together with several other views [30] for cases with suspected abnormalities. In this work, only the processed FFDms with expert’s annotated ground-truth are used, resulting in a total of 2145 cases with cancers.

There are several breast abnormalities in the OMI-DB dataset, such as masses, calcifications, architectural distortions, focal asymmetries, or combinations of the above. Since the focus of this paper is on the detection of masses or mass like abnormalities, mammograms with calcifications only are not considered, while architectural distortions and focal asymmetries are included in the dataset. The categorisation of the OMI-DB dataset based on BI-RADS [31] ratings and mass conspicuity is shown in Fig. 1 (only the biopsy proven BI-RADS ratings are considered). The OMI-H and OMI-G dataset contained, respectively 2042 and 103 positive cases, with abnormalities in either one of the mammography views (CC and MLO), and 842 and 104 normal cases, i.e. without any abnormalities.

2.2. INbreast

The INbreast [16] public dataset is composed of FFDms acquired using Siemens MammoNovation mammography system (Siemens Healthineers, Erlangen, Germany). The FFDms are acquired from 115 cases with CC and MLO mammography views, leading to a total of 410 FFDms available in DICOM format. From these, a total of 116 masses can be found in 107 mammograms from 50 cases. The masses are divided into benign (BI-RADS $\in \{2,3\}$) and malignant (BI-RADS $\in \{4,5,6\}$).

2.3. Data preparation

Details of the dataset in this work is shown in Table 1, where Cases refers to patients with different mammographic views and Images is the number of processed images. The positive images are obtained from the cases with abnormalities, while all the negative images are obtained from those with no abnormalities (normal cases). The extra negative images of the cases with abnormalities, in which there is only one breast or one view, are not considered as they are the clinical history of the same patient.

The OMI-H dataset is divided into training, validation, and testing sets in the ratio of 70%, 10%, and 20% respectively. In the OMI-G and INbreast datasets, a 5-fold cross-validation strategy is used to test all the mammograms in the datasets. The division of images is done on a patient basis such that all the mammograms from an individual case belongs exclusively in either training or testing set. Note that some images in the OMI-H and OMI-G datasets (referred to as unknown), whose BI-RADS ratings are not clearly stated, are also annotated as masses by the radiologists. Therefore, we used these cases for training the model using the label as mass.

Table 1

Description of dataset used in this work: Pos refers to positives (masses) and Neg refers to negatives (non-mass).

Dataset	OMI-H	OMI-G	INbreast
Resolution μm	70, 60	100	70
Image sizes (pixels)	3328 \times 2560 4096 \times 3328	2294 \times 1914 3062 \times 2394	3328 \times 2560 3328 \times 4084
Cases	Pos Neg	2042 842	103 104
Images	Pos Neg	3770 3475	50 65
Mass	Benign Malign Unknown	485 3048 684	11 143 52
Image distribution	Training Validation Testing	5316 585 1344	5-folds cross-validation cross-validation

2.4. Data pre-processing

The original mammograms in the datasets are high-resolution images ($\sim 60\text{--}100\ \mu\text{m}$) with sizes in the range of $\sim 2000\text{--}4000$ pixels. To focus the processing in the areas with intensity information (non-background pixels), mammograms are cropped identifying the breast area bounding box. Subsequently, and in line with other works in the literature, mammograms are down-sampled to $200\ \mu\text{m}$ for computational and memory limitations. Moreover, the images in the OMI-G and INbreast datasets have very different contrast compared to those in the OMI-H dataset, so image normalisation is performed on the images in these datasets. In the OMI-G, a window width (WW) and window center (WC) normalisation are applied [32]. The WW and WC information is obtained from the DICOM header of the mammograms. For images in the INbreast dataset, an adjustment is made on the windows of the pixel levels. The images are normalised (similar to Ribli et al. [20]) and intensity re-scaled to 8 bits.

3. Faster R-CNN

The Faster R-CNN model has been widely used for detecting objects particularly in natural image datasets, e.g. PASCAL VOC [34], MS-COCO [35], etc. Advancements in the field of object detection are often built on the success of region proposal methods [36] and R-CNNs [4]. For example, in 2015, Ren et al. [5] introduced region proposal network (RPN), which takes an image as the input and outputs a set of rectangular boxes used to detect and localise objects in the image. These rectangular boxes are characterised by an ‘‘objectness’’ score, which is a measure of closeness of the detected object to a certain object class. The Faster R-CNN [5] uses RPN, along with the Fast R-CNN [37] model, to accelerate the training and testing processes, and to improve performance. In the Faster R-CNN paradigm, the problem of object detection is considered as both a regression and classification problem. Additionally, during training of the Faster-RCNN model, a class balance is performed using a mini-batch obtained from a single image, such that positive and negative anchors are obtained in a 1:1 ratio, and the loss function of the mini-batch is then computed [5].

In the Faster R-CNN API [33], a collection of pre-trained detection models is provided, which includes: 21 models pre-trained on MS-COCO dataset [35], 1 model on KITTI dataset [38], 6 models on Open Images [39], and 2 models on iNaturalist species [40]. Table 2 presents some of the available pre-trained models along with their speed. In this paper, the Faster R-CNN model (InceptionV2 [41] as feature extractor) pre-trained on the MS-COCO dataset is used. This selection is based on the performance of Faster R-CNN with different backbone models and datasets in terms of trade-off between speed and average precision.

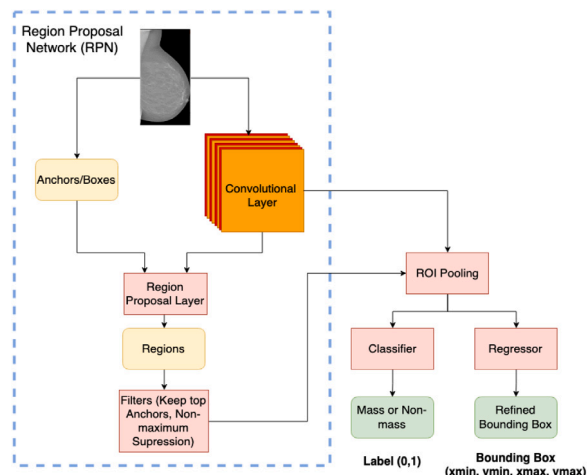


Fig. 2. Flow chart of the Faster R-CNN, showing the region proposal network (RPN) and the overall pipeline. The convolutional layer of the InceptionV2 is used as feature extractor, which shares the feature map with the RPN, generating the region proposals. The classification and regression problem are solved on these proposals to generate the final bounding box information.

4. Methodology

In the first step, FFDM is used as the input to the Faster R-CNN model. This is forwarded through the InceptionV2 [41] based CNN model to produce the feature map. In the second step, an RPN is created using the extracted features of the CNN and is trained to detect and localise masses on the mammograms. A window of size $n \times n$ slides over the feature map and outputs a feature vector linked to two fully convolutional (FC) layers, i.e. box-regression layer (reg) and box-classification layer (cls). Thereafter, the anchors or bounding boxes are created to generate region proposals of varying shapes and sizes, and are given an objectness score signifying how accurately they are enclosing a mass on the mammogram. The highest scoring anchors are then passed to the second stage of the network. Here, a classification and regression problem is solved to accurately detect the presence of masses, and simultaneously refine the coordinates of the anchors to precisely detect them. In the last step, the best predictions are obtained by using non-maximum suppression on the detected overlapping objects, resulting in the final detected bounding box with confidence probability representing how close it represents a mass (Fig. 2).

In this paper, the Faster R-CNN model proposed by Ren et al. [5] 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. Note that all the computations are performed on a Linux workstation with 12 CPU cores (3.4 GHz) and an NVIDIA TitanX Pascal GPU with 12GB memory.

4.1. Training and hyperparameter tuning

The Faster R-CNN model is implemented within the Tensorflow object detection API [33]. The input to the model is in the form of Tfrecords (containing the mammograms along with class definitions, bounding box coordinates, etc.). The training is performed using the hyperparameter ‘‘keep aspect ratio’’ with the maximum height and width of the mammograms in the entire dataset. As we have a large dataset, only horizontal flipping is applied as data augmentation during the training process. In terms of the optimiser, the *Stochastic gradient descent method with momentum* [42] is used, with a momentum value of 0.9. During training, the learning rate is heuristically decreased in steps after every 25,000 iterations and continued until 200,000 iterations.

Table 2
Some of the pre-trained Faster R-CNN models available in [33].

Model	Training dataset	Speed (ms)	mean average precision
Faster_RCINN_Inception_v2	MS COCO	58	28
Faster_RCINN_Resnet50	MS COCO	89	30
Faster_RCINN_Resnet101	MS COCO	106	32
Faster_RCINN_Inception_Resnet_v2_atrous_v2	Open Images	727	37
Faster_RCINN_Resnet101_fgvc	iNaturalist	395	58

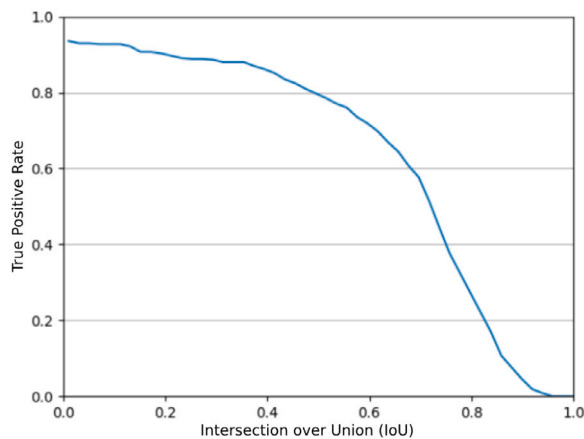


Fig. 3. Comparison of TPR vs IoU for the detection of masses in OMI-H testing dataset.

Additionally, as all the mammograms in the dataset are cropped to the breast profile resulting in the mammograms with different pixel dimensions, a batch size of 1 is used for training.

The anchors are created with a base size of 128 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), resulting in a total of 15 anchors at a defined pixel location. This selection of anchor scales and aspect ratios is done based on the average size distribution of the ground-truth masses in the entire dataset. At the second stage, the detections are processed with non-maximum suppression using a threshold of 0.05, resulting in non-overlapping bounding box predictions to represent masses in the mammogram. This is done to avoid overlapping detection boxes, as in mammograms it is less likely to have overlapping masses [20].

4.2. Transfer learning

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 [43]. Azizpour et al. [44] suggests 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 being transferred. In this paper, the transfer learning methodology is used to fine-tune the Faster R-CNN model pre-trained on a large mammography dataset (OMI-H) to detect masses in small mammography datasets (OMI-G and INbreast) obtained using different scanners.

4.3. Evaluation metric

In this work, the objectness score is obtained as an output of the network, which is then used as the confidence probability to construct the confusion matrix. To assess the classification and detection performance of the proposed 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-receiver operating curve (FROC). The qualitative assessment is made using the confusion matrix to compute the sensitivity, specificity, and F1-score of the proposed framework as:

$$\begin{aligned} \text{sensitivity} &= \frac{TP}{TP + FN}, \\ \text{specificity} &= \frac{TN}{TN + FP}, \\ \text{F1-score} &= \frac{2 \times TP}{2 \times TP + FN + FP} \end{aligned} \quad (1)$$

where, TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives per mammograms, respectively. For the detection framework, a mass is considered to be detected as a true positive (TP), if the intersection over union (IoU), defining the overlapping area between the predicted box, and the ground truth (GT) box, is greater than a pre-defined threshold. This threshold is obtained by evaluating the detection performance at different IoUs (Fig. 3), where the TPR is plotted along with the detection IoU. It can be seen that TPR is almost constant for $0.0 < \text{IoU} \leq 0.1$, slightly reduces (0.93–0.88) for $0.1 < \text{IoU} \leq 0.3$, and starts to fall sharply for $\text{IoU} \geq 0.6$. Therefore, in further sections, an IoU of 0.1 is used for all testing sets to compare the predicted and GT results (as used by Ribli et al. [20]).

In cases of multiple masses in a single mammogram, the confusion matrix is computed for the whole mammogram. The IoU is calculated separately for each mass, and if $\text{IoU} \geq 0.1$ for one of the predicted bounding boxes, the masses are considered to be detected (true positive, TP). Masses that remained undetected are considered as false negatives (FNs), while all remaining prediction boxes with $\text{IoU} < 0.1$ are considered false positives (FPs) for the mammogram. The area (A_z) under the receiver operative curve (ROC) is also used to evaluate the mass classification results. The accuracy of the detection results is assessed using the FROC curve, which is plotted as the function of True Positive Rate (TPR) versus the False Positives per Image (FPI).

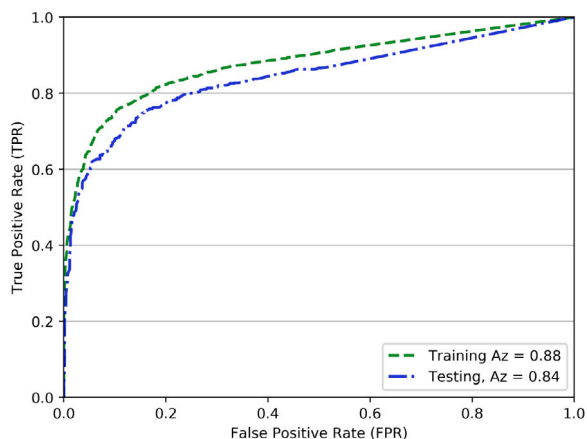
5. Experimental results

5.1. Mass detection on OMI-H

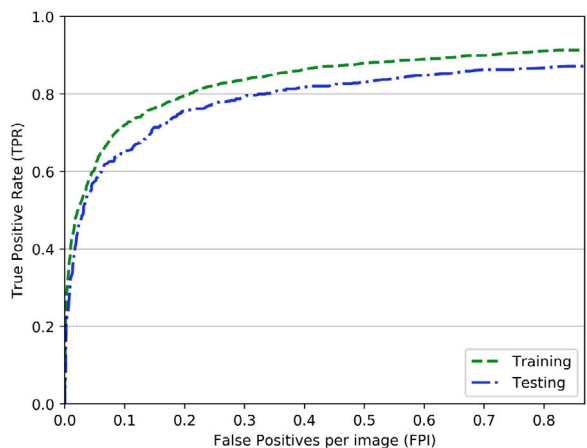
In this section, a domain transfer is performed between natural images and mammograms to detect masses in the latter. This is done by adapting the Faster R-CNN model, pre-trained on a large dataset of natural images (MS-COCO), to detect masses in the mammograms. The pre-trained Faster R-CNN model is fine-tuned using a training set of 5316 processed mammograms acquired using the Hologic scanner (OMI-H). Note that the training for lesion and non-lesion is done using a balanced dataset (Table 1). The model obtained after training is tested on 1344 (testing set) mammograms, and the predictions on each mammogram compared against the available ground truth (GT) bounding box annotations.

The model's performance is evaluated using the ROC curve (Fig. 4(a)) achieving $A_z = 0.88$ and $A_z = 0.84$ on the training and testing data respectively. Additionally, in the plotted FROC curve (Fig. 4(b)), it can be seen that the best result of $\text{TPR} = 0.91$ at 0.82 FPI is obtained on the training data, and $\text{TPR} = 0.87$ at 0.84 FPI is obtained on the testing data. Furthermore, F1-score of 0.734 has been obtained for the testing data.

In the field of breast cancer imaging, the limited availability of large annotated datasets has been a limiting factor for the success of deep



(a)



(b)

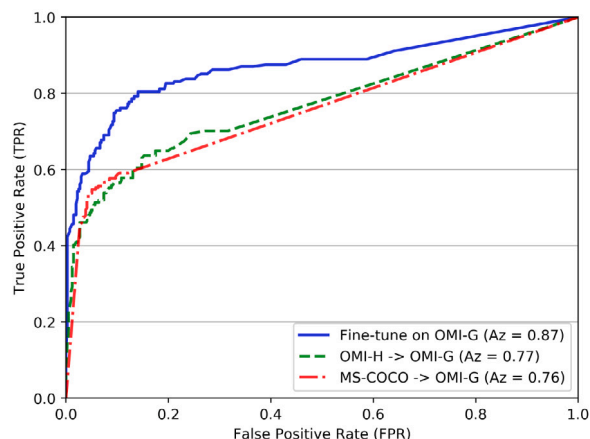
Fig. 4. Mass detection on OMI-H training and testing dataset, (a) ROC curve, (b) FROC curve.

learning methodologies. In the following sections, the transfer learning methodology is used to fine-tune the Faster R-CNN model pre-trained on the OMI-H dataset to detect masses in small mammography datasets: OMI-G and INbreast.

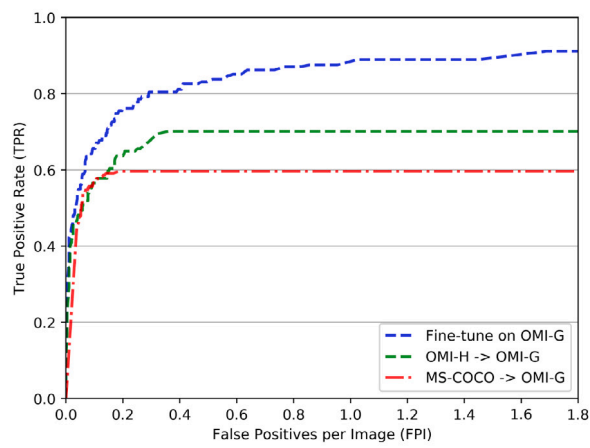
5.2. Mass detection in OMI-G

To train the Faster-RCNN model using a small mammography dataset OMI-G, an analysis is performed. To achieve this, the OMI-G 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 OMI-G dataset. The trained model showed reduced performance achieving an $Az = 0.76$ and TPR of 0.60 at 0.20 FPI. Thereafter, the model trained on the OMI-H dataset is used (without fine-tuning) to detect and localise masses in the test set of the OMI-G dataset, resulting in TPR of 0.70 at 0.43 FPI and $Az = 0.77$.

Lastly, fine-tuning is used to adapt the feature domain of the Faster R-CNN model, trained on OMI-H dataset, to detect masses in the OMI-G dataset using a 5-fold cross-validation strategy, resulting in TPR of 0.91 ± 0.06 at 1.70 FPI, $Az = 0.87$, and F1-score of 0.80. The per-fold confusion matrix results are presented in Table 3, while the overall detection results are summarised in Table 4. Additionally, the ROC and FROC curves from these experiments are shown in Figs. 5(a) and 5(b) respectively.



(a)



(b)

Fig. 5. Mass detection on OMI-G testing dataset, (a) ROC curve, (b) FROC curve. The results are shown for OMI-H model tested directly on OMI-G and OMI-H model fine-tuned on OMI-G.

Table 3

Per-fold confusion matrix results: OMI-H→OMI-G (confidence probability ≥ 0.20).

	TP	FP	FN	TN
Fold 1	22	22	5	67
Fold 2	18	24	2	71
Fold 3	20	22	5	63
Fold 4	17	12	8	80
Fold 5	12	9	3	71

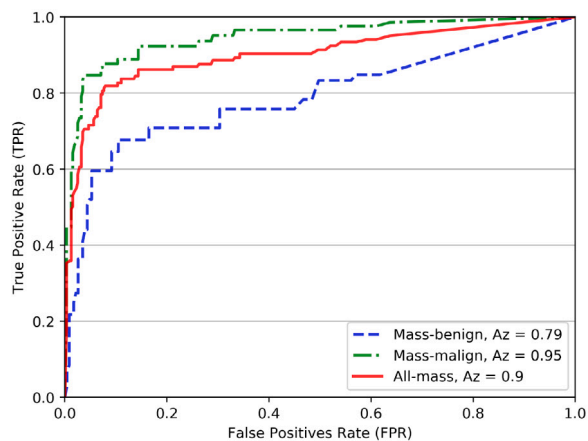
Table 4

Mass detection results: OMI-H → OMI-G.

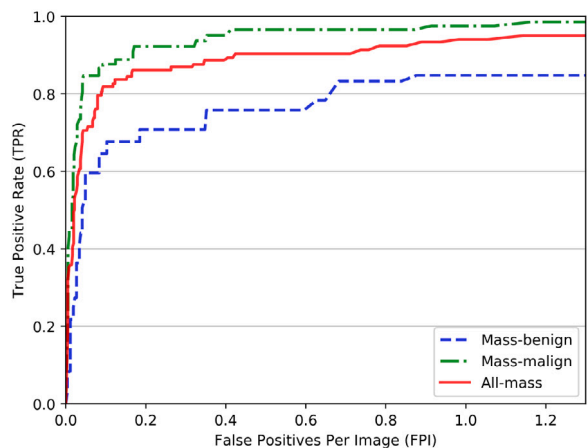
Model	Trained on OMI-G	Trained on OMI-H	Fine tuned on OMI-G
Sensitivity	0.58	0.57	0.76
Specificity	0.88	0.88	0.88
Az	0.76	0.77	0.87
TPR at FPI	0.59 ± 0.04 at 0.13	0.63 at 0.20 0.70 at 0.43	0.74 ± 0.06 at 0.20 0.91 ± 0.06 at 1.69

5.3. Mass detection in INbreast

The Faster R-CNN model is also used to detect masses in the INbreast public dataset. While the proposed detection framework is used



(a)



(b)

Fig. 6. Mass detection on INbreast testing dataset, (a) ROC curve (b) FROC curve. The results are shown for malignant, benign and all masses.

to detect masses in general, the performance is also analysed based on malignant and benign lesions.

Firstly, the performance of the Faster R-CNN model, pre-trained on the OMI-H dataset is analysed to directly detect the masses in the full INbreast dataset (without fine-tuning), resulting in $Az = 0.89$ for the malignant masses (BI-RADS $\in \{4,5,6\}$), and $Az = 0.67$ for the benign masses (BI-RADS $\in \{2,3\}$). Moreover, the FROC analysis resulted in a TPR of 0.87 at 0.32 FPI for the malignant masses, and 0.55 at 0.32 FPI for the benign masses.

Secondly, the model, trained on the OMI-H dataset, is fine-tuned on the INbreast dataset. A 5-fold cross-validation strategy is used to analyse the entire INbreast dataset. In contrast to OMI-DB, there are fewer positive cases (50 numbers) in the INbreast dataset. Therefore, an augmented training dataset is created using very small rotation angles, i.e. $0^\circ, \pm 5^\circ, \pm 10^\circ$. This data augmentation is used only for the purpose of training the model.

The detection performance is analysed using the ROC and FROC curves as shown in Fig. 6(a), and Fig. 6(b) respectively. The detection framework of the fine-tuned model (OMI-H \rightarrow INbreast) resulted in a sensitivity of 0.99 ± 0.03 at 1.17 FPI ($Az = 0.95$) with F1-score of 0.86 for the malignant masses, and 0.85 ± 0.08 at 1.0 FPI ($Az = 0.79$) with F1-score of 0.74 for the benign masses. The per-fold confusion matrix results are presented in Table 5, while a summary of the overall detection results is presented in Table 6.

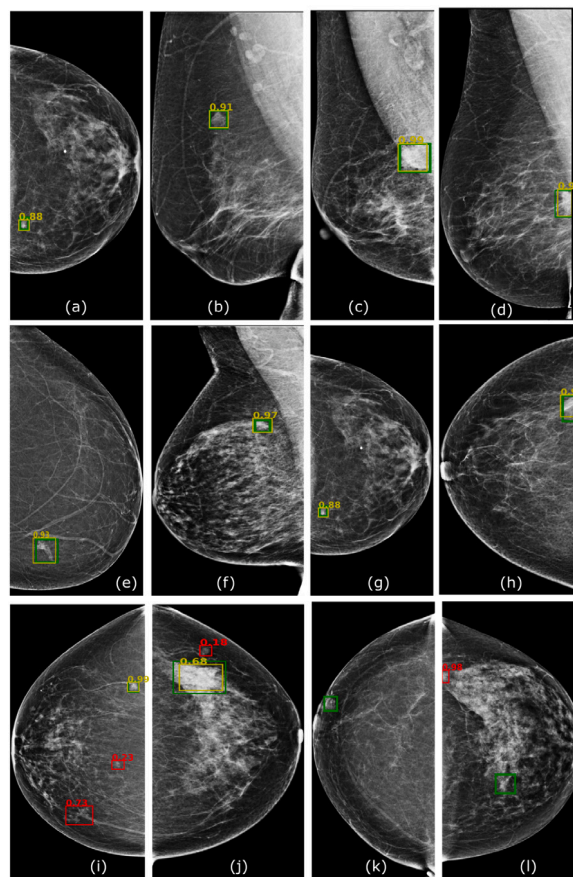


Fig. 7. Mass detection results in OMI-H dataset, (a–h) demonstrate detections with high objectness score, (i, j) shows some detections with FPs, and (k, l) shows undetected masses (green: GT box, yellow and red: detection box). The numbers shown in images corresponds to the confidence of being mass.

Table 5
Per-fold confusion matrix results:
OMI-H \rightarrow INbreast (confidence probability ≥ 0.20).

	TP	FP	FN	TN
Fold 1	23	29	2	40
Fold 2	18	14	4	53
Fold 3	27	27	4	40
Fold 4	16	24	3	41
Fold 5	18	51	0	40

5.4. Qualitative results

In Fig. 7, examples of mass detection results are visualised on the mammograms in the OMI-H datasets. Several prediction results are shown: the top two rows show mammograms with precisely predicted masses; ground truth annotations are displayed in green, and the predicted boxes with their confidence scores are in yellow. In Fig. 7(i, j) FP detections are shown (red) along with TP, Fig. 7(k, l) which show undetected masses. Several mass detection results in the OMI-G and INbreast datasets are visualised in Fig. 8 and Fig. 9 respectively. Figs. 8(a–h) and 9(a–d) show a single mass detection result at the precise position with a high confidence score in each mammogram, and Fig. 9(e, h) shows the detection of several masses in the same mammogram. Some undetected masses are shown in Figs. 8(i) and 9(k, l).

Table 6
Mass detection results: OMI-H→INbreast.

Model	Trained on OMI-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
Az	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

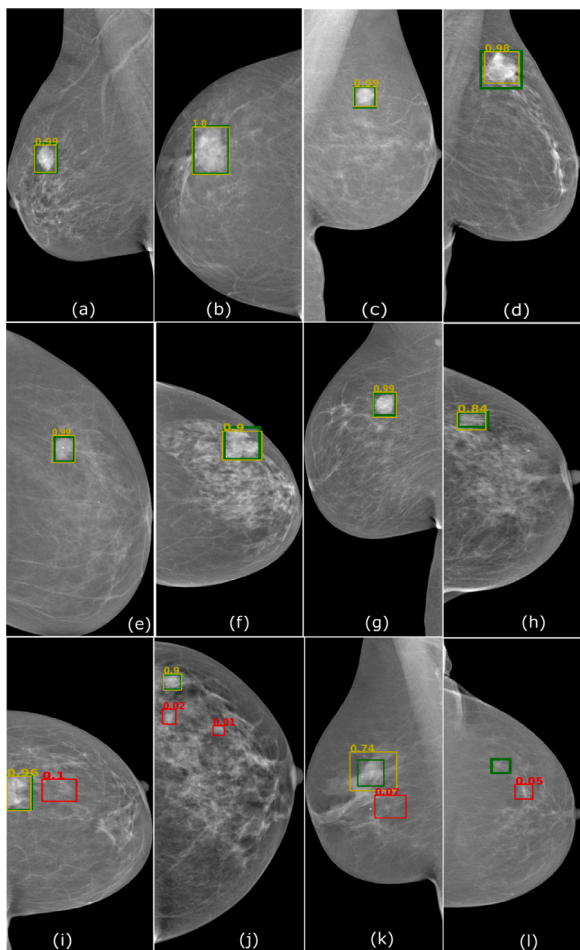


Fig. 8. Mass detection results in OMI-G dataset, (a–h) shows detections with high confidence score, (i–k) shows detections with FPs, and (l) shows undetected mass (green: GT box, yellow and red: detection box). The numbers shown in images corresponds to the confidence of being mass.

6. Discussions

In this paper, it has been shown that the Faster R-CNN model, pre-trained on an entirely different dataset of natural images, can be fine-tuned to efficiently detect masses in whole mammograms. It has also been shown that enhanced performance can be obtained when the Faster R-CNN model is trained on a large database of mammograms (OMI-H), and then fine-tuned using the mammograms in smaller databases (OMI-G and INbreast).

In the OMI-H dataset, analysis of the undetected masses was also performed based on the conspicuity. It was found that 8.7% of the total obvious masses were not detected (51 out of 601). Moreover, 5 out of 20 very subtle masses (25%), 26 out of 139 subtle masses (18.7%), and 3 out of 4 occult masses (75%) remained undetected using the proposed framework. The mass detection was also analysed based

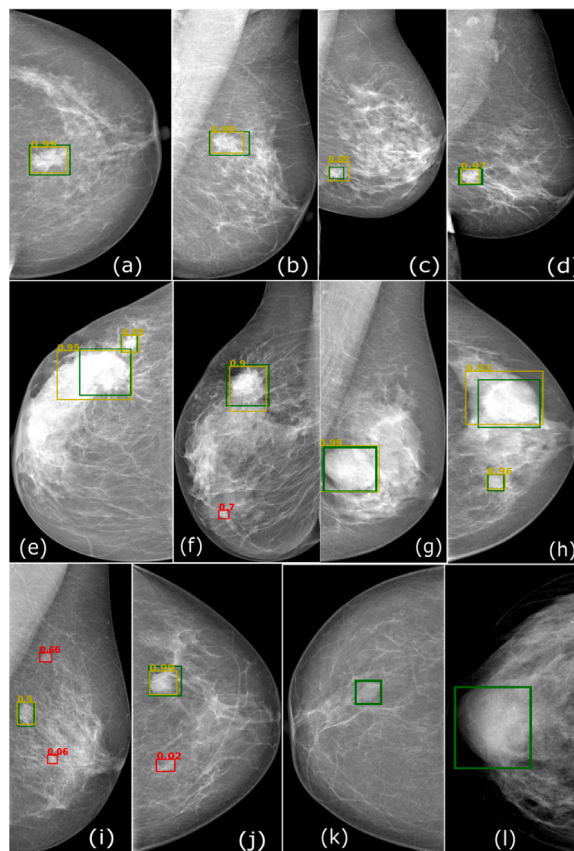


Fig. 9. Mass detection results on INbreast dataset, (a–d,g) shows detections with high confidence score, (e, h) show multiple detections in the same mammogram, (f, i, j) shows detections with FPs, and (k, l) shows undetected masses (green: GT box, yellow and red: detection box). The numbers shown in images corresponds to the confidence of being mass.

on the malignancy, i.e. malignant (BI-RADS ∈ {4,5,6}) and benign (BI-RADS ∈ {2,3}). It was found that the detection framework was able to detect malignant masses with a sensitivity of 0.86 at 1.2 FPI and 0.73 at 0.30 FPI, while for benign masses, the model resulted in a sensitivity of 0.81 at 0.72 FPI and 0.65 at 0.30 FPI.

In terms of transfer learning from different scanners, the images in the OMI-G dataset had low contrast compared to the OMI-H dataset. Thus, normalisation of the images in these datasets was performed and the results compared in Table 4. Compared to the results obtained for the original mammograms in the OMI-G dataset ($Az = 0.77$), higher performance was obtained for the normalised images ($Az = 0.87$) ($p < 0.005$). This can be justified as the Faster R-CNN model was pre-trained on the OMI-H images, which are similar in contrast to the normalised images in the OMI-G dataset. This contrast enhancement benefits the fine-tuning process, and as expected, the system detected approximately 83% of masses in the OMI-G dataset with 0.43 FPIs, and with 1.7 FPIs, more than 90% of masses are detected.

In principle, it would be feasible to mix the two datasets (OMI-H and OMI-G) for training even though sensor sizes (field of view

and pixel sizes) have some differences. This can be done in order to increase the training size (especially for OMI-G), given that the input images are converted to the same pixel size (i.e. 200 μm). Although we performed this experiment, testing results for both OMI-H and OMI-G were worse compared to training with the same vendor dataset. This could be explained by the fact that, although performing an intensity normalisation step, images from different vendors still show differences in dynamic range, local contrast, and signal to noise ratio, amongst others, mainly due to their imaging characteristics and post-processing.

The availability of a large dataset to train the Faster R-CNN model has been one of the limitations of the framework proposed in the current work. This is evident from comparing the mass detection performance on the OMI-G dataset using the two approaches: (i) Faster R-CNN trained directly on OMI-G ($A_z = 0.76$), and (ii) Faster R-CNN firstly trained on OMI-H, and then fine-tuned on OMI-G ($A_z = 0.87$) ($p < 0.005$). To the best of our knowledge, this is the first study presenting the results on OMI-DB. Although a comparison with any other works could not be established for these datasets, a comparison to other works using large mammography dataset was established. For instance, Kooi et al. [14] obtained an A_z of 0.93 on an internal database of $\sim 45,000$ FFDMs; Morrel et al. [23] obtained an A_z of 0.87 on 13,000 images from Group Health. In this work, we obtained an A_z of 0.90 for the malignant masses in the OMI-H dataset.

The benefits of transfer learning were evaluated by fine-tuning the OMI-DB trained model to detect masses in a public dataset (INbreast) and comparisons were done with state-of-the-art methods (Table 7). It is evident from the results that the performance of the proposed framework is comparable or higher than existing methods.

The mass detection model of Akselrod-Ballin et al. [45] obtained a TPR of 0.93 at 0.56 FPI on a subset of the INbreast dataset (100 mass images), compared to a TPR of 0.87 at 0.3 FPI obtained in this work (410 images). Although, the detection performance of the proposed framework is slightly lower, there is a difference in the number of images used and thus a one-to-one comparison is difficult to establish. Dhungel et al. [8] proposed a framework consisting of a cascade of deep learning methods. This was aimed at reducing the false positive detections to subsequently improve the precision of the bounding box predictions, obtaining TPR of 0.90 ± 0.02 at 1.3 FPI for the INbreast dataset. In this paper, we obtained a TPR of 0.95 ± 0.03 at 1.14 FPI using a much simpler object detection framework.

Ribli et al. [20] used the Faster R-CNN model with VGG16 and trained the model on the INbreast dataset consisting not only the masses, but also the calcifications. They evaluated the performance on malignant masses and calcifications to obtain a TPR of 0.90 at 0.3 FPI. In this paper, we obtained comparatively higher detection results with a TPR of 0.92 at 0.3 FPI. However, a direct comparison is difficult to establish because of differences in the selection of mammograms with malignant masses.

Jung et al. [21] used a one-stage object detection model (RetinaNet) and trained it using an in-house dataset containing malignant masses along with INbreast dataset. It has been shown in previous works [22], that compared to Faster R-CNN, higher performance has been achieved by using RetinaNet for detecting objects in natural images. However, we show that higher performance is achieved for detecting masses in mammograms using the Faster R-CNN model. This can be linked to the fact that a much larger dataset (10 times) of FFDMs, containing malignant masses, has been used to train the Faster R-CNN model compared to 222 FFDMs used for training the RetinaNet model in [21].

Al-masni et al. [9] used the YOLO object detection model to propose a CAD system to handle the detection and classification simultaneously. The authors showed enhanced accuracies when using an augmented DDSM database created by rotating each mammogram multiple times. In other work, Al-antari et al. [10] proposed an automated framework to detect, segment and classify masses in a single framework. As per our understanding, the training and testing sets in both the works are

non-exclusive, and thus the one-to-one comparison with the results in this study cannot be established.

In one of our previous work [28], traditional CNN-based methods were employed to obtain patch level predictions which were used to detect masses in whole mammograms by reconstructing the whole image (from patches) using sliding window approach. The framework obtained similar (or better) results than state-of-the-art methods on the INbreast dataset (TPR of 0.98 ± 0.02 at 1.67 FPI). However, the high computational cost associated with the image reconstruction step was a limiting factor for clinical use.

The inference time per mammogram of some of the methods are shown in Table 7. Although, the computational environment used for different methods could be different, this gives an overall idea regarding the applicability of the methodology in real-time. The inference time per image of the proposed Faster-RCNN method is 4 s per image using a GPU with 12 GB memory, which is slightly higher compared to single-stage methods like YOLO, RetinaNet, etc. (Table 7).

In this work, the results are shown in mammograms obtained from three different scanners, i.e. Hologic (OMI-H), GE (OMI-G) and Siemens (INbreast), demonstrating the adaptability of the presented framework to detect masses in the images obtained using different mammography acquisition systems. Furthermore, in the OMI-G dataset, a sensitivity of 0.75 is achieved with a specificity of 0.90, and a TPR of 0.75 at 0.20 FPI. For malignant masses in the public dataset INbreast, a sensitivity of 0.87 is achieved with a specificity of 0.90, and TPR of 0.87 at 0.1 FPI. These results suggest the proposed framework is able to detect masses with a reduced number of false positives and false negatives, demonstrating its potential use as a part of a CAD system to aid radiologists in the detection of breast cancers. As noted previously, the focus of this study is currently on mass detection, so images containing calcifications were not considered for the analysis. This will be investigated further in future studies.

7. Conclusions and future work

In this work, the implementation of a Faster R-CNN model for detecting masses in a large-scale dataset of breast mammograms was presented. It was shown that the Faster R-CNN model, pre-trained on an entirely different dataset of natural images, could be adapted to efficiently detect masses in whole mammograms. It was also shown that enhanced performance could be obtained when the Faster R-CNN model was trained on a large database of mammograms and fine-tuned to adapt it for other mammograms in smaller databases (OMI-G and INbreast). Compared to other works in the literature, the proposed mass detection framework showed improved performance in terms of higher sensitivity and lower FPI.

The proposed framework has the potential to be used within the clinical environment as it takes the whole mammogram as the input and outputs the suspected masses within the mammogram. Moreover, the presented framework has been tested to detect masses in two different small mammography datasets obtained using different mammography acquisition systems. This demonstrates the potential of the detection framework for analysing mammograms from different systems, which is a particular requirement for successful deployment in different clinical environments.

In terms of future work, we intend to extend this to classify masses into benign and malignant. Moreover, it would be interesting to perform an observer study involving experienced radiologists (at least 2–3) and comparing their standalone performance results against the mass detection tool presented here. Additionally, an investigation would be performed to adapt the presented method to detect masses in 3D volumes such as Automated Breast Ultrasound and Digital Breast Tomosynthesis [50], which are currently being adopted in clinical practice.

Table 7
Comparison between proposed framework and the published results.

Methods	TPR ($\mu \pm \sigma$) at FPI	# Dataset	# Images	Inference time per image (s)
Kozegar et al. [46]	0.87 at 3.67 0.91 at 4.8	INbreast mini-MIAS	107 330	108
Akselrod et al. [45]	0.93 at 0.56 0.90 at 1.0	INbreast Private	100 750	5
Shen et al. [47]	0.8788 at 0.5 0.9479 at 2.0	INbreast Internal	32 282	N/A
Anitha et al. [48]	0.935 at 0.62 0.925 at 1.06	mini-MIAS DDSM	170 300	N/A
Brake et al. [49]	0.55 at 0.10	DDSM	772 (Malignant)	N/A
Dhungel et al. [8]	0.90 \pm 0.02 at 1.3	INbreast	410	39
Ribli et al. [20]	0.90 at 0.30	INbreast	Malignant only	N/A
Jung et al. [21]	0.94 \pm 0.05 at 1.30	INbreast	410	1.8
Al-masni et al. [10]	N/A	INbreast	816	3
Agarwal et al. [28]	0.98 \pm 0.02 at 1.67	INbreast	410	35
Proposed framework	0.93 at 0.78	OMI-H	1219 (Malignant)	3
	0.91 \pm 0.06 at 1.70	OMI-G	601	
	0.92 \pm 0.08 at 0.30	INbreast	Malignant	
	0.85 \pm 0.08 at 1.0		Benign	
	0.95 \pm 0.03 at 1.14		410	

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is partially supported by SMARTER project funded by the Ministry of Economy and Competitiveness of Spain, under project reference DPI2015-68442-R, and the ICEBERG project (Ref. RTI2018-096333-B-I00) funded by the Ministry of Science, Innovation and Universities. R. Agarwal is funded by the support of the Secretariat of Universities and Research, Ministry of Economy and Knowledge, Government of Catalonia Ref. ECO/1794/2015 FIDGR-2016. The authors would like to thanks Dr. Lucy Warren from Royal Surrey County Hospital (UK) for her help and support in understanding the OMI-DB. The authors gratefully acknowledge the support of the NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research. We would like to thank Kim's English Corner (<https://kimsenglishcorner.com>) for proofreading.

References

- [1] B. Stewart, 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, F. Bray, Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012, *Eur. J. Cancer* 49 (6) (2013) 1374–1403.
- [3] Eurostat health statistics, 2019, Online, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cancer_statistics_-_specific_cancers. (Last Accessed 30 May 2019).
- [4] R. Girshick, J. Donahue, T. Darrell, 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.
- [5] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6) (2017) 1137–1149.
- [6] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8) (2013) 1915–1929.
- [7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, 2015, pp. 1–9.
- [8] N. Dhungel, G. Carneiro, A.P. Bradley, A deep learning approach for the analysis of masses in mammograms with minimal user intervention, *Med. Image Anal.* 37 (2017) 114–128.
- [9] 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, T.-S. Kim, Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLOs-based CAD system, *Comput. Methods Programs Biomed.* 157 (2018) 85–94.
- [10] M.A. Al-antari, M.A. Al-masni, M.-T. Choi, S.-M. Han, T.-S. Kim, A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification, *Int. J. Med. Inf.* 117 (2018) 44–54.
- [11] A. Oliver, J. Freixenet, J. Marti, E. Perez, J. Pont, E.R. Denton, R. Zwiggelaar, A review of automatic mass detection and segmentation in mammographic images, *Med. Image Anal.* 14 (2) (2010) 87–110.
- [12] M.L. Giger, N. Karssemeijer, J.A. Schnabel, Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer, *Annu. Rev. Biomed. Eng.* 15 (2013) 327–357.
- [13] J. Arevalo, F.A. González, R. Ramos-Pollán, J.L. Oliveira, M.A.G. Lopez, Representation learning for mammography mass lesion classification with convolutional neural networks, *Comput. Methods Programs Biomed.* 127 (2016) 248–257.
- [14] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C.I. Sánchez, R. Mann, A. den Heeten, N. Karssemeijer, Large scale deep learning for computer aided detection of mammographic lesions, *Med. Image Anal.* 35 (2017) 303–312.
- [15] M. Heath, K. Bowyer, D. Kopans, R. Moore, P. Kegelmeyer, The digital database for screening mammography, *Digit. Mammogr.* (2000) 431–434.
- [16] I.C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M.J. ao Cardoso, J.S. Cardoso, Inbreast: toward a full-field digital mammographic database, *Academic Radiol.* 19 (2) (2012) 236–248.
- [17] S. Vesal, S.M. Patil, N. Ravikumar, 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.
- [18] M. Goyal, N. Reeves, S. Rajbhandari, M.H. Yap, Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices, *IEEE J. Biomed. Health Inf.* (2018).
- [19] A. Akselrod-Ballin, L. Karlinsky, S. Alpert, S. Hashoul, R. Ben-Ari, E. Barkan, A CNN based method for automatic mass detection and classification in mammograms, *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* 7 (3) (2019) 242–249.
- [20] D. Ribli, A. Horváth, Z. Unger, P. Pollner, I. Csabai, Detecting and classifying lesions in mammograms with Deep Learning, *Sci. Rep.* 8 (1) (2018) 4165.
- [21] H. Jung, B. Kim, I. Lee, M. Yoo, J. Lee, S. Ham, O. Woo, J. Kang, Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network, *PLoS One* 13 (9) (2018) e0203355.
- [22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
- [23] S. Morrell, Z. Wojna, C.S. Khoo, S. Ourselin, 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.
- [24] J. Dai, Y. Li, K. He, J. Sun, R-FCN: Object detection via region-based fully convolutional networks, in: Advances in Neural Information Processing Systems, 2016, pp. 379–387.

- [25] M.D. Halling-Brown, P. Looney, M. Patel, L. Warren, A. Mackenzie, K. Young, The oncology medical image database (OMI-DB), in: *Medical Imaging 2014: PACS and Imaging Informatics: Next Generation and Innovations*, Vol. 9039, International Society for Optics and Photonics, 2014, p. 903906.
- [26] Dream challenge: The digital mammography Dream challenge, 2017, Online, <https://www.synapse.org/#!Synapse:syn4224222/wiki/401743>. (Last Accessed 26 December 2017).
- [27] J. Redmon, S. Divvala, R. Girshick, 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.
- [28] R. Agarwal, O. Diaz, X. Lladó, M.H. Yap, R. Martí, Automatic mass detection in mammograms using deep convolutional neural networks, *J. Med. Imaging* 6 (3) (2019).
- [29] C. Gu, J. Lim, P. Arbelaez, J. Malik, Recognition using regions, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [30] Mammography views, 2020, Online, http://dicom.nema.org/medical/Dicom/2014c/output/chtml/part03/sect_C.8.11.7.html. (Last Accessed 03 March 2020).
- [31] E. Sickles, C. D'Orsi, L. Bassett, C. Appleton, W. Berg, E. Burnside, et al., ACR BI-RADS[®] mammography, in: *ACR BI-RADS[®] Atlas, Breast Imaging Reporting and Data System*, Vol. 5, American College of Radiology, Reston, VA, 2013.
- [32] DICOM, 2019, Online, http://dicom.nema.org/medical/dicom/current/output/chtml/part03/sect_C.11.2.html. (Last Accessed 05 June 2019).
- [33] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, et al., Speed/accuracy trade-offs for modern convolutional object detectors, in: *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2017.
- [34] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, *Int. J. Comput. Vis.* 88 (2) (2010) 303–338.
- [35] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, *Microsoft COCO: Common Objects in Context*, Springer International Publishing, 2014, pp. 740–755.
- [36] J.R. Uijlings, K.E. Van De Sande, T. Gevers, A.W. Smeulders, Selective search for object recognition, *Int. J. Comput. Vis.* 104 (2) (2013) 154–171.
- [37] R. Girshick, Fast R-CNN, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.
- [38] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? The KITTI vision benchmark suite, in: *Conference on Computer Vision and Pattern Recognition*.
- [39] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Mallocci, T. Duerig, V. Ferrari, The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale, 2018.
- [40] G.V. Horn, O.M. Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, S. Belongie, The inaturalist species classification and detection dataset, in: *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2018.
- [41] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, 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.
- [42] N. Qian, On the momentum term in gradient descent learning algorithms, *Neural Netw.* 12 (1) (1999) 145–151.
- [43] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks? in: *Advances in Neural Information Processing Systems*, NIPS, 2014, pp. 3320–3328.
- [44] H. Azizpour, A.S. Razavian, J. Sullivan, A. Maki, S. Carlsson, From generic to specific deep representations for visual recognition, in: *CVPRW DeepVision Workshop*, June 11, 2015, Boston, MA, USA, IEEE, 2015.
- [45] A. Akselrod-Ballin, L. Karlinsky, A. Hazan, R. Bakalo, A.B. Horesh, Y. Shoshan, 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.
- [46] E. Kozegar, M. Soryani, B. Minaei, I. Domingues, Assessment of a novel mass detection algorithm in mammograms, *J. Cancer Res. Ther.* 9 (4) (2013) 592.
- [47] R. Shen, J. Yao, K. Yan, K. Tian, C. Jiang, K. Zhou, Unsupervised domain adaptation with adversarial learning for mass detection in mammogram, *Neurocomputing* (2020).
- [48] J. Anitha, J.D. Peter, S.I.A. Pandian, A dual stage adaptive thresholding (DuSAT) for automatic mass detection in mammograms, *Comput. Methods Programs Biomed.* 138 (2017) 93–104.
- [49] G. te Brake, N. Karssemeijer, J.H. Hendriks, An automatic method to discriminate malignant masses from normal tissue in digital mammograms, *Phys. Med. Biol.* 45 (10) (2000) 2843.
- [50] M. Fan, Y. Li, S. Zheng, W. Peng, W. Tang, L. Li, Computer-aided detection of mass in digital breast tomosynthesis using a faster region-based convolutional neural network, *Methods* 166 (2019) 103–111.