Breast Ultrasound Region of Interest Detection and Lesion Localisation

Moi Hoon Yap[a,*], Manu Goyal[a], Fatima Osman[b], Robert Martí[c], Erika Denton[d], Arne Juette[d], Reyer Zwiggelaar[e]

[a] *Department of Computing and Mathematics, Manchester Metropolitan University, UK*
[b] *Department of Computer Science, Sudan University of Science and Technology, Sudan.*
[c] *Computer Vision and Robotics Institute, University of Girona, Spain*
[d] *Nolfolk and Norwich University Hospital Foundation Trust, Norwich, UK*
[e] *Department of Computer Science, Aberystwyth University, UK.*

Abstract

In current breast ultrasound Computer Aided Diagnosis systems, the radiologist preselects a region of interest (ROI) as an input for computerized breast ultrasound image analysis. This task is time consuming and there is inconsistency among human experts. Researchers attempting to automate the process of obtaining the ROIs have been relying on image processing and conventional machine learning methods. We propose the use of a deep learning method for breast ultrasound ROI detection and lesion localisation. We use the most accurate object detection deep learning framework – Faster-RCNN with Inception-ResNet-v2 – as our deep learning network. Due to the lack of datasets, we use transfer learning and propose a new 3-channel artificial RGB method to improve the overall performance. We evaluate and compare the performance of our proposed methods on two datasets (namely, Dataset A and Dataset B), i.e. within individual datasets and composite dataset. We report the lesion detection results with two types of analysis: 1) *detected point* (centre of the segmented region or the detected bounding box) and 2) Intersection over Union (IoU). Our results demonstrate that the proposed methods achieved comparable results on *detected point* but with notable improvement on IoU. In addition, our proposed 3-channel artificial RGB method improves the *recall* of Dataset A. Finally, we outline some

*Corresponding author
Email address: `m.yap@mmu.ac.uk` (Moi Hoon Yap)

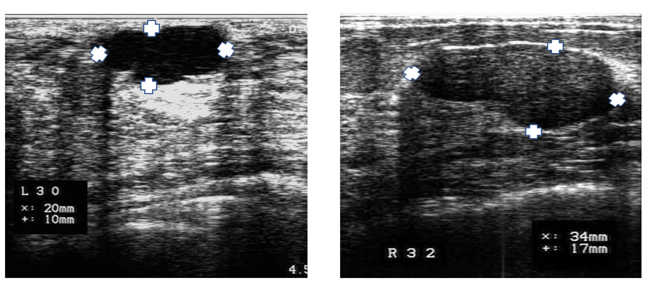
future directions for the research.

Keywords:
Breast ultrasound, breast cancer, object detection, region of interests

1. Introduction

Breast cancer is a common disease for women and is considered to be the second leading cause of death worldwide [1]. According to Breast Cancer Now [2], breast cancer is the most common cancer in the UK. Ultrasound is the complementary modality to the standard imaging method (two view mammography) in breast cancer diagnosis [3, 4]. It is the most widely used in clinical practice [5] compared to other alternatives such as tomosynthesis and magnetic resonance imaging. Due to the fact that early detection plays a main role in avoiding breast cancer deaths and increases the proportion of healing and recovery, there has been increasing interest in using ultrasound to aid in the early detection of breast cancers over the past few years [6, 7].

In Breast Ultrasound (BUS), radiologists are trained in interpreting the sonographic features [8]. In current practice, the clinician scans the breast and takes static images. The radiologist will assess and annotate the BUS images. Computer Aided Diagnosis (CAD) systems are then can be used as a "second reader" for computerized medical imaging analysis [9]. These systems are based on the assumption that the radiologist detects an abnormality and preselects a region of interest (ROI). Figure 1 shows BUS images with manual pre-selected ROIs marked with '+' and 'x'.



(a) (b)

Figure 1: Examples of BUS images with manual pre-selected ROIs marked with '+' for the upper and lower points for the lesion, and 'x' for the leftmost and rightmost points of the lesion. Please note that the annotations were embossed for better visualisation.

Previous work attempted to automate the process of ROIs selection [10, 11, 12, 13]. These methods were based on multi-stage image processing and/or machine learning approaches. Deep learning has gained popularity in biomedical image analysis and has achieved good results in classification [6, 14] and BUS semantic segmentation [15]. Yap et al. [7] compared the performance of lesions detection algorithms and showed that deep learning approaches are more accurate and robust across datasets. However, the limitations of their work were: 1) they detected the lesions by using segmentation approaches but not an object detection approach; and 2) they evaluated the performance based on *detected point* (centre of the segmented region) [7], not the overlap of the regions.

According to state-of-the-art BUS lesion detection [6, 16], a ROI is defined as a bounding box circumscribing the lesion. This paper focuses on the automatic detection of such ROIs. We propose the use of the Faster-RCNN Inception-ResNet-v2 approach [17] for BUS lesion detection. The key contributions are:

1. We automate the ROI detection using a popular deep learning approach, this is the first attempt in automation of BUS ROI detection using Faster-RCNN Inception-ResNet-v2.
2. We propose two approaches to overcome the issue of lack of BUS data. First we apply a transfer learning approach and then we propose a new 3-channel artificial RGB method to improve the quality of results.
3. We evaluate and compare the performance of our proposed method on two datasets - within individual datasets and composite dataset. As existing approaches do not focus on ROI bounding box detection, we compare the performance of our proposed methods with FCN-AlexNet.

2. Related Work

In current practice, the clinical expert manually locates rectangular sub-images [18, 19] to locate ROIs on BUS images. However, in large-scale studies, this step is time-consuming. Hence, researchers [20, 21, 10] have developed algorithms to locate the ROIs automatically. Within fully automated ROI detection, there are two types of ROI: 1) ROI as an initial contour of the lesion [20, 21, 22, 23]; and 2) ROI as a rectangle region containing both lesion and some background information [10, 12]. In this section, we review research on both ROI definitions.

Table 1: A Comparison of Dataset A and Dataset B.

Comparison	**Dataset A**	**Dataset B**
Capture Devices	B&K Medical Panther 2002 and B&K Medical Hawk 2012	Siemens ACUSON Sequoia C512 system
Transducer	8-12 MHz linear array transducer	8.5 MHz 17L5 HD linear array transducer
Year	2001	2012
Number of Images	306	163
Image size	377×396	760×570

In 1998, based on a single feature called the radial gradient index (RGI), Kupinski et al. [20] developed a novel lesion segmentation technique. Using gray-level information, and prior knowledge of the shape of typical mass lesions, a series of image partitions were created and the partition that maximised the RGI was selected. The method was tested on a database of biopsy-proven, malignant lesions. According to their results [20], the RGI segmentation algorithm correctly segmented 92% of the lesions. Although the work of Kupinski et al. [20] assessed the RGI filter in mammograms, it was applied to BUS images in 2002 by Drukker et al. [21], where the use of RGI filtering technique was investigated for automated lesion detection in BUS. Using a database of 757 images from 400 patients, lesion candidates were segmented from the background by maximising an average radial gradient index for regions grown from the *detected point*. Initial RGI filtering achieved a sensitivity of 87% at 0.76 false-positive detections [21].

(a) Extreme points (b) Binary mask (c) Bounding box

Figure 2: The ground truth format conversion of BUS datasets: (a) original extreme points; (a) original segmentation ground truth in binary mask form provided by Yap et al. [7]; and (c) conversion to bounding box as the ground truth for ROI detection and localisation.

In 2008, Yap et al. [10] proposed a novel approach for boundary detection of ROI in BUS images. In the preprocessing step, histogram equalization was applied, followed by a combination of nonlinear diffusion and linear filtering. Further to this hybrid filtering stage, the visually distinct areas of the BUS image were analysed using multifractals. In the final stage, region growing based segmentation was applied to partition the filtered BUS image using different threshold values. According to the assumption of Kupinski et al. [20], selection of the lesion was made by choosing the partition with the highest RGI. The work indicated that multifractal analysis could be useful for enhancing boundary detection in ultrasound images.

For the detection of masses, Ikedo et al. [24] used a feature based on the edge directions in each slice, and a method for subtracting between slices. In order to detect edges, a Canny edge detector was applied and morphology was used to classify the detected edges into two groups: near-vertical edges or near-horizontal edges. Subsequently, the near-vertical edges were used as cues, then using the segmented and the low-density regions, they were able to segment the located positions by a watershed algorithm, and mass candidate regions were detected. Finally, for the distribution between masses and false positives (FPs), rule-based schemes and a quadratic discriminant analysis were applied in order to remove FPs. Aiming to improve the screening performance and efficiency, the proposed scheme achieved sensitivity of 80.6% with 3.8 FPs per breast image.

A fully automated segmentation method was proposed in 2012 by Shan et al. [12]. Two main findings were introduced: an efficient ROI generation method and new features to characterise lesion boundaries were proposed. In order to develop an automatic ROI generation method, two steps were used, the first step was the automatic seed point selection and the second was a region growing step. Region growing was considered to be fast and simple, although its accuracy was not high, it was serving the purpose as it roughly located the lesion rather than finding the accurate boundary of it. Further, they combined traditional intensity-and-texture features and two proposed lesion features (phase in max-energy orientation and radial distance) were used to detect lesions by a trained artificial neural network. On a database of 120 images, the method improved the true positive (TP) rate from 84.9% to 92.8%, the similarity rate from 79.0% to 83.1% and reduced the FP rate from 14.1% to 12.0%.

In order to detect lesions in breast US images, with no need for any kind of human interaction or supervision, Pons et al. [25] proposed a feasibility study by adapting a generic object detection technique, called Deformable Part Models (DPM). They provided an assessment of this methodology to lesion detection by applying it for the first time to US images, using a dataset of 100 images, all from different patients (50 were healthy tissue regions, 18 were malignant lesions, 32 were benign lesions). According to results for lesion detection, they showed the feasibility of their proposal and they achieved a sensitivity of 82% with 0.51 false-positive detections per image and an A_z value of 0.96.

Although research to date has demonstrated the feasibility to automate the ROI detection by using computer algorithms, like in similar medical image analysis research, there are some common issues:

1. Research was conducted within a single institution or hospital; code and datasets were not shared. Therefore, the research is not reproducible, and less straight forward to compare.
2. The use of performance metrics has not been consistent, i.e. some used FP rate, while others used FP per image; some reported sensitivity and specificity, while others used *recall* and *precision*.
3. The methods were mostly based on image processing and conventional machine learning. Although some researchers [7, 15, 6] have been working actively in deep learning for classification and segmentation, the use of deep learning for ROI detection in BUS is yet to be fully explored.

We address these issues by proposing the use of a popular deep learning method for ROI detection on two publicly available datasets, and we report the results with a variety of performance metrics. If the manuscript is accepted for publication, the codes will be made available on github.

3. Methodology

This section discusses the BUS datasets, the preparation of the ground truth labeling, the proposed ROI detection method (based on transfer learning, the 3-channel Artificial RGB image method and a Faster-RCNN approach) and the performance metrics for the ROI detection results.

3.1. Datasets and Ground Truth

In general, ultrasound images are complex because of data composition, which can be described in terms of speckle information. Upon visual inspection, ultrasound images could be described as speckle noise that varies between bright and dark degrees of grayscale. The two datasets (henceforth, Dataset A and Dataset B) that we used in this paper were obtained from a recent publication by Yap et al. [7]. They are referred to as Dataset A and Dataset B and Table 1 compares the two datasets. The 306 images in Dataset A are from 2001. Although Dataset A might not be a representative of clinical practice, it is still interesting to test the robustness of machine learning algorithms on different image resolutions. The 163 images in Dataset B are from 2012 and have a higher image resolution. To standardize the image resolution for our experiments, we have resized the images to 500×375. For a detailed description and to download Dataset B, please refer to [7].

The ground truths provided in the BUS datasets are in the form of binary masks of the lesions or with extreme points, as illustrated in Fig. 2(a). From these extreme points, we generated rectangle bounding boxes around the binary masks for ROI localisation. Fig. 2(b) illustrates an example of a bounding box overlaid on the original BUS image. This is a mandatory step as the bounding boxes are commonly used in computer vision as the ground truth labels to train the object detection algorithms.

3.2. Transfer Learning

To obtain good performance, current state-of-the-art deep learning methods require large-scale datasets to train the model [26]. In natural images, large-scale datasets exist such as ImageNet [27] and the MS-COCO dataset [28]. ImageNet [27] consists of more than 1.5 million images for the classification of 1000 pre-defined classes [27] and the MS-COCO dataset [28] consists of 328,000 images with 91 common object categories. To use these pre-trained models for our proposed BUS ROI lesion detection framework, we convert the original grayscale BUS images to 3-channel images (I) by concatenating three single channel grayscale images (I_g) from the BUS datasets, as shown in Equation 1.

$$I = Concat(I_g, I_g, I_g) \tag{1}$$

where I is a 3-channel converted image from the concatenation of three original grayscale images (I_g).

Transfer learning is a popular technique in deep learning to overcome data deficiency, where we can choose to transfer the features from a few convolutional layers (partial transfer learning) or from all layers (full transfer learning) of a pre-trained model. For our proposed framework, we implemented

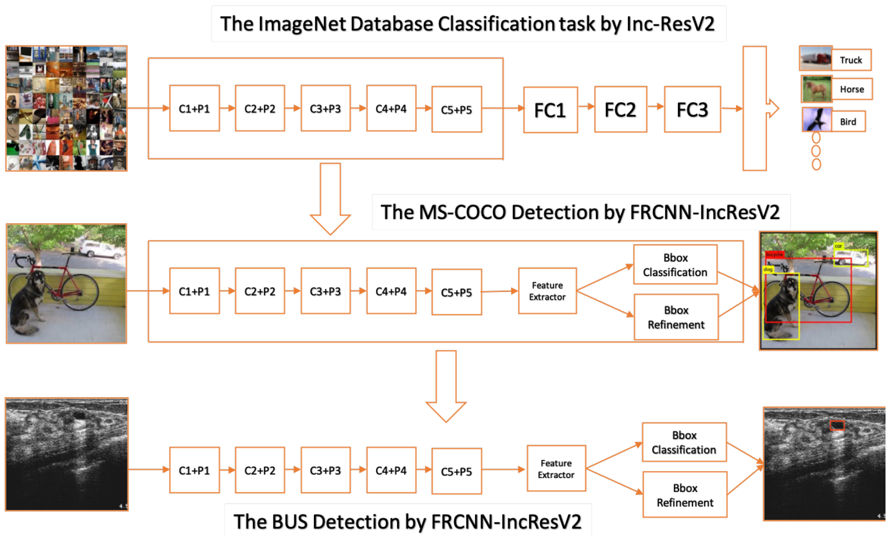


Figure 3: Overview of two-tier transfer learning used for ROI detection and localisation of BUS lesions.

two-tier transfer learning [29]. Firstly we used partial transfer learning by transferring the features only from the convolutional layers trained on the most significant classification challenge dataset - ImageNet. Then, we used full transfer learning from a model trained on MS-COCO object localisation dataset as shown in Fig. 3.

3.3. 3-channel Artificial RGB Image Method

In standard data augmentation techniques, the number of training images is increased with different image manipulation algorithms, including rotation, flipping and image filtering. Data augmentation has shown to be effective in improving the performance of deep learning algorithms. However, it has increased the time and memory requirements in training the algorithms. We propose a new 3-channel artificial RGB image method by concatenating the original image with two post-processed images. With this proposed technique, we maintain the number of training images, i.e. rather than concatenating the three grayscale images, we used two filtered images to concatenate with the grayscale image. The proposed 3-channel artificial RGB image (I_a) is produced by concatenating a single channel grayscale image (I_g), the sharpened image (I_s) and the contrast enhanced image (I_c), as shown in Equation 2.

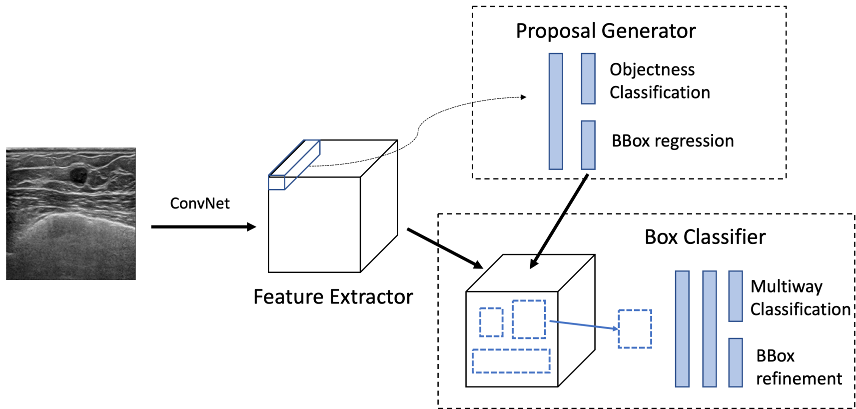


Figure 4: Overview of the proposed architecture (redrawn from [30]) for BUS experiments. The Proposal Generator generates Bounding Box (BBox) from the feature maps. The refinement and classification of BBox proposals are attained by Inception-ResNet-v2 to obtain the best accuracy of BBox.

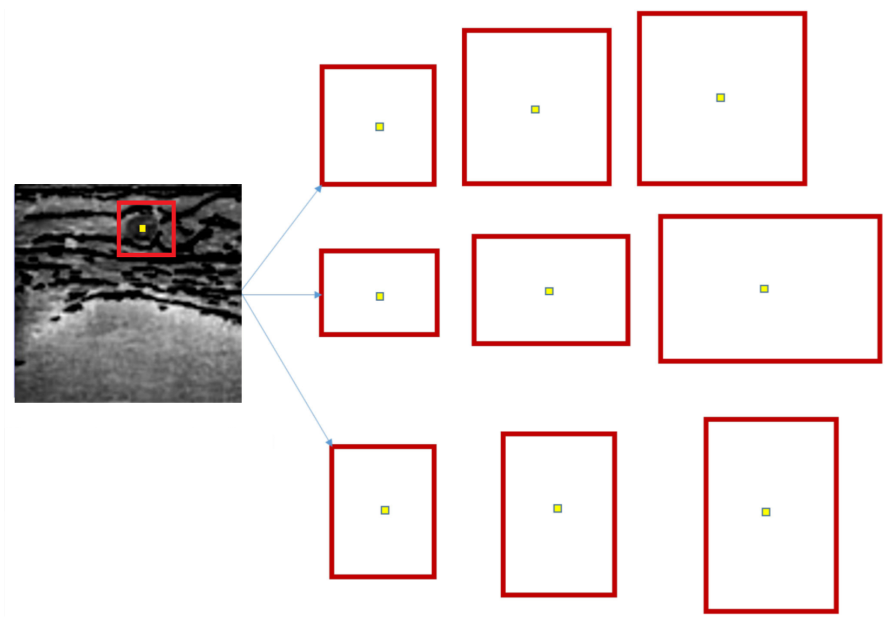


Figure 5: Nine different anchors are generated for a single point of the feature map.

$$I_a = Concat(I_g, I_s, I_c) \tag{2}$$

3.4. Faster-RCNN Inception-ResNet-v2 approach

Faster-RCNN Inception-ResNet-v2 is one of the most accurate state-of-the-art models for object localisation [30]. It has been successfully implemented, e.g. in person detection [28] and diabetic foot ulcers localisation [31]. In the earlier version of the Region Proposal Network (RPN), the first step is to generate region proposals by selective search, then classify and detect the object based on a Convolutional Neural Network (CNN) framework. The core design of the Faster-RCNN was similar to the Region-based CNN, i.e. hypothesise object regions based on the feature maps and then classify them using the similar CNN. The benefit of Inception-ResNet-v2 [17] is it combined the optimization benefits conferred by residual connections with the computation efficiency of Inception units. Figure 4 illustrates the architecture for Faster-RCNN [32] with Inception-ResNet-v2 approach [17] . The architecture of Faster-RCNN consists of three stages:

- First Stage: A pre-trained CNN (Inception-ResNet-v2) was used to extract the convolutional feature map of BUS images from the last convolutional layer for proposal generator (Second Stage) and BBox classification and regression (Third Stage).

- Second Stage: The proposal generator is used to find a predefined number of bounding box (BBox) proposals, may contain a lesion. Anchors are fixed bounding boxes that are placed throughout the image with different sizes (64px, 128px, 256px) and ratios (0.5, 1, 1.5) to find lesions in the BUS image as shown in Fig. 5. Then, two layers (objectness classification layer and BBox regression layer) are used to find the "objectness score" for these anchors to have a good set of BBox proposals. For this stage, as BUS images have a very limited number of lesions (mostly one lesion per image), we set the value of a number of proposals to 100.

- Third Stage: Finally, these BBox proposals (from the Second Stage) are then passed through a pre-trained CNN in the next step to extract features for each proposal. The ROI pooling layer is used to produce fixed-size feature maps from non-uniform inputs of proposals by performing a max pooling operation. These features are finally used by the

Box Classifier (classification and BBox refinement layers) to refine and classify the proposals, which obtains the final accurate BBox regions. We only chose BBox regions with confidence equal to 90% or higher for final evaluation.

3.5. Performance Metrics

We used four popular performance metrics i.e. *Precision*, *Recall*, *F1-Score* and *False Positives per Image (FPI)* for the evaluation of BUS detection and localisation. The state-of-the-art BUS lesion detection research used *detected point* criterion [7]. However, the measurement based on the centre of detected bounding box or segmented region can be misleading. To overcome this issue, we use "overlap criterion" as an Intersection over Union (IoU) greater than 0.5 [33]. The IoU is defined by equation 3.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{3}$$

In the context of medical image analysis, IoU is known as the Jaccard Similarity Index or Jaccard Index. Based on IoU as the criteria, we calculate the following parameters:

1. *True Positives (TP)* defined as Bounding Boxes (BBox) that have IoU greater than 0.5 with the BB of the ground truth (GT).
2. *False Positives (FP)* defined as BBox that have IoU less than 0.5 with GT and also, the duplicate BB that have IoU with a GT that has already been detected.
3. *True Negatives (TN)*: In BUS datasets, all the images contain at least one lesion. This is due to current practice that the clinician will only save ultrasound images with a lesion. Hence, there were no normal images and we can not obtain *TN*.
4. *False Negatives (FN)* were calculated if there is no detection of the BBox produced by the algorithm.

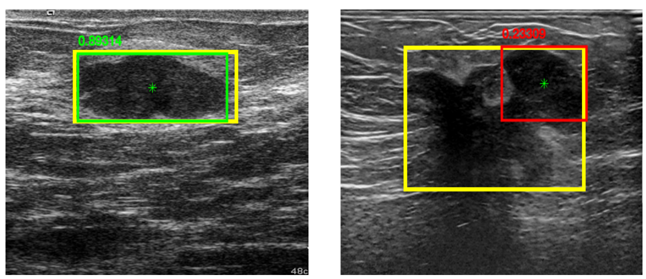
The *Precision* was calculated by total number of correct BBox i.e. *TP* divided by the total number of ground truth i.e. *TP* and *FP*, as shown in equation 4. The *Recall* was the total number of correct detected bounding boxes (i.e. *TP*) divided by total number of detected bounding boxes (i.e. *TP*) and *FN*, as in equation 5. The last evaluation metric was the *F1-Score*, which was the harmonic average of *Precision* and *Recall* (see equation 6). The *F1-Score* is also known as Dice Coefficient Index in medical image analysis.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 - Score = \frac{2 \times (Recall \times Precision)}{Recall + Precision} \quad (6)$$

To compare with state-of-the-art methods, we also report our results as in Yap et al. [7], i.e. detection is considered as a TP if the detection point (centre of the detected bounding box) is placed within the ground truth bounding box of an expert radiologist. Otherwise, it was considered to be a *FP*. Figure 6 compares the differences between two criteria, where IoU is more reliable in reporting the results.



(a) (b)

Figure 6: The yellow box indicates ground truth, the green '*' indicated the *detected point*, the green bounding box indicates true detection and the red bounding box indicates false detection: (a) this is an example of both *detected point* and IoU achieved agreement with a True Positive; and (b) this is an example where even the detected region is at the top right corner, the *detected point* calculated as true detection but the IoU has a more strict measurement and categorised it as a false detection.

3.6. Implementation

For consistency, we have evaluated all the methods using 5-fold cross-validation on 3-channel grayscale datasets and 3-channel artificial RGB datasets. For the composite dataset (combination of dataset A and dataset B), this was not totally random as we needed to ensure the training set distributions consisted of both datasets. For the benchmark algorithm, we used the Caffe framework [34] to implement the transfer learning FCN-AlexNet. We repeated the experiment using similar settings as in [7], where the model was trained using stochastic gradient descent with a learning rate of 0.001, 60 epochs with a dropout rate of 33%. To convert the segmentation results produced by FCN-AlexNet, we used the similar method in converting the binary masks to ground truth bounding boxes, where the coordinates of the left most pixel, the top most pixel, the right most pixel and the bottom most pixel are used to form the bounding box (as illustrated in Fig. 2).

For the implementation of the Faster-RCNN Inception-ResNet-v2 approach (henceforth, FRCNN), we used the original parameters as in [17], with the learning rate of 0.001. We observed the models converged at 100 epochs. Our experiments were run on a GPU machine with the following configurations: (1) Hardware: CPU - Intel i76700@4.00 Ghz, GPU - NVIDIA TITAN X 12 GB, RAM - 32 GB DDR4 (2) Deep Learning framework: Tensor-flow.

4. Result and Discussion

We performed thorough evaluation within and between the datasets. We evaluated the results based on 5-fold cross validation on single datasets (solely on Dataset A and Dataset B) and composite dataset (A+B). We reported the results of the individual dataset in the composite dataset experiment, which was (A+B) on A and (A+B) on B. We discuss the results in two detection methods, i.e. *detected point* and *IoU*. Then we perform visual comparison of the results.

4.1. Evaluation based on detected point

Table 2 shows the overall FRCNN results based on *detected point*. From the results of Yap et al. [7], the transfer learning FCN-AlexNet (henceforth, FCN-AlexNet) [35] outperformed Radial Gradient Index Filtering [21], Multifractal Filtering [10], Rule-based Region Ranking [12], Deformable Part Models [13], and two deep learning techniques (U-Net [36] and Patched-based LeNet [37]). To compare the performance of FRCNN on BUS lesion detection, we used FCN-AlexNet as the benchmark algorithm.

4.1.1. Within dataset analysis

We observed all the methods were obtaining high *recall* and *precision* when evaluated based on *detected point*. Although the performance of FR-CNN obtained the best results in this setting, the *recall* for FCN-AlexNet is comparable. Overall, FRCNN achieved the best *F1-Score* but FRCNN with

Table 2: Comparison of performance metrics based on *detected point* for ROI detection in BUS dataset. FRCNN is Faster-RCNN Inception-ResNet-v2 on concatenated grayscale BUS images whereas FRCNN (RGB) is Faster-RCNN Inception-ResNet-v2 on 3-channel artificial RGB BUS images. FCN-AlexNet represents transfer learning FCN-AlexNet. Bold indicates the best result for each category and underline indicates the best result for the Dataset.

Dataset	Method	*Recall*	*Precision*	*F1-Score*	*FPI*
A	FCN-AlexNet	0.9388	0.8365	0.8847	0.1961
	FRCNN	0.9236	<u>**0.9408**</u>	<u>**0.9321**</u>	<u>**0.0621**</u>
	FRCNN (RGB)	<u>**0.9572**</u>	0.9020	0.9288	0.1111
B	FCN-AlexNet	0.9080	0.8605	0.8836	0.1472
	FRCNN	**0.9141**	<u>**0.9371**</u>	<u>**0.9255**</u>	<u>**0.0614**</u>
	FRCNN (RGB)	0.8589	0.8861	0.8723	0.1104
(A+B) on A	FCN-AlexNet	0.9450	0.8351	0.8867	0.1994
	FRCNN	**0.9480**	**0.8857**	**0.9158**	**0.1307**
	FRCNN (RGB)	0.8746	0.8338	0.8537	0.1863
(A+B) on B	FCN-AlexNet	0.9325	0.7917	0.8563	0.2454
	FRCNN	<u>**0.9632**</u>	**0.8441**	**0.8997**	**0.1779**
	FRCNN (RGB)	0.8344	0.7953	0.8144	0.2147

3-channel artificial RGB images achieved the best *recall* of 0.9572 for Dataset A. For dataset B, the *recall* of FRCNN marginally improved FCN-AlexNet but FCN-AlexNet produced more *FPs*. Overall, FRCNN achieved the best *F1-Score* with 0.9321 and 0.9255 on Dataset A and Dataset B, respectively.

4.1.2. Composite dataset analysis

When compared the composite results, FCN-AlexNet and FRCNN improved in terms of *recall* but with poorer performance in *precision*. These were due to the methods detecting more regions when trained on two datasets with different modalities. However, for FRCNN with the 3-channel artificial RGB technique, the results were less satisfactory for all the metrics. This has demonstrated that even though 3-channel artificial RGB images proved to improve the *recall* of Dataset A, which can be caused by introduction of

Table 3: Comparison of performance metrics based on *IoU* for ROI detection in BUS dataset. FRCNN is Faster-RCNN Inception-ResNet-v2 on concatenated grayscale BUS images whereas FRCNN (RGB) is Faster-RCNN Inception-ResNet-v2 on 3-channel artificial RGB BUS images. FCN-AlexNet represents transfer learning FCN-AlexNet. Bold indicates the best result for each category and underline indicates the best result for the Dataset. *STD* represents standard deviation.

Dataset	Method	*IoU* (*mean*±*STD*)	*Recall*	*Precision*	*F1-Score*	*FPI*
A	FCN-AlexNet	0.7800±0.1069	0.8624	0.7684	0.8127	0.2778
	FRCNN	0.8447±0.0946	0.8838	**0.9003**	0.8920	**0.1046**
	FRCNN (RGB)	**0.8535±0.0888**	**0.9358**	0.8818	**0.9080**	0.1340
B	FCN-AlexNet	0.7145±0.1123	0.6749	0.6395	0.6567	0.3804
	FRCNN	**0.8363±0.0863**	**0.8773**	**0.8994**	**0.8882**	**0.0982**
	FRCNN (RGB)	0.8254±0.0919	0.8221	0.8481	0.8349	0.1472
(A+B) on A	FCN-AlexNet	0.7837±0.1066	0.8716	0.7703	0.8178	0.2778
	FRCNN	0.8496±0.0904	**0.9205**	**0.8600**	**0.8892**	**0.1601**
	FRCNN (RGB)	**0.8532±0.0860**	0.7584	0.7230	0.7403	0.3105
(A+B) on B	FCN-AlexNet	0.7537±0.1151	0.7485	0.6354	0.6873	0.4295
	FRCNN	0.8395±0.0930	0.8896	**0.7796**	**0.8310**	**0.2515**
	FRCNN (RGB)	**0.8399±0.0896**	0.7485	0.7135	0.7305	0.3006

noisy data and hence become less robust across datasets. Overall, FRCNN is the most robust method across different datasets.

Since the measurement solely based on the *detected point* of the bounding box could be misleading, the following section reports the results based on overlap criterion – *IoU*.

4.2. Evaluation based on IoU

Table 3 summarises the results based on the overlap criterion of *IoU* greater than 0.5. We report the results based on single datasets and composite dataset.

4.2.1. Within dataset analysis

Since the overlap criterion followed a more strict rule, we observed all the performance metrics were poorer when compared to the *detected point*. Particularly the performance of FCN-AlexNet notably decreased for all the evaluation. Interestingly, the FRCNN with 3-channel artificial RGB images worked the best on Dataset A with *recall* of 0.9358 and *F1-Score* of 0.9080. However, FRCNN achieved the best results on Dataset B with *recall* of 0.8773, *precision* of 0.8994 and *F1-Score* of 0.8882. We observed the FR-CNN with 3-channel artificial RGB images has achieved the best *IoU* when evaluated on Dataset A.

4.2.2. Composite dataset analysis

Similar to the results on *detected point*, FRCNN was the most robust algorithm for the composite dataset analysis across all the performance metrics. FCN-AlexNet has shown marginal improvement when compared to the within dataset analysis. However, FRCNN with 3-channel artificial RGB images has deteriorated with very poor results. This has demonstrated that even though 3-channel artificial RGB images proved to improve the *recall* and *F1-Score* of Dataset A, it is not robust across datasets. A similar finding shows FRCNN is more robust across different datasets when measured by *IoU*. To further demonstrate the result, the following section reports qualitative analysis.

4.3. Visual Comparison

Figure 7 visually compares the results of the proposed methods and FCN-AlexNet. The yellow boxes indicate ground truth, the green boxes indicate *TP* when *IoU* greater than 0.5, the red boxes indicate *FP*, the green '*' indicates *TP* for *detected point*, the red '*' indicate *FP* for *detected point*. The first row of Figure 7 shows a best case for all the algorithms, the second row shows the detected lesion by FRCNN but not FCN-AlexNet, and the third row illustrates a complex case where all the algorithms achieved different results. It is interesting to observe that FCN-AlexNet has a *TP* for *detected point* but a *FP* for *IoU* criterion.

4.4. Summary

From the results, we summarise our observations as follow:

- The overall performance of FCN-AlexNet was better on the composite dataset. This implies that it is more suitable for larger heterogeneous scale of dataset.

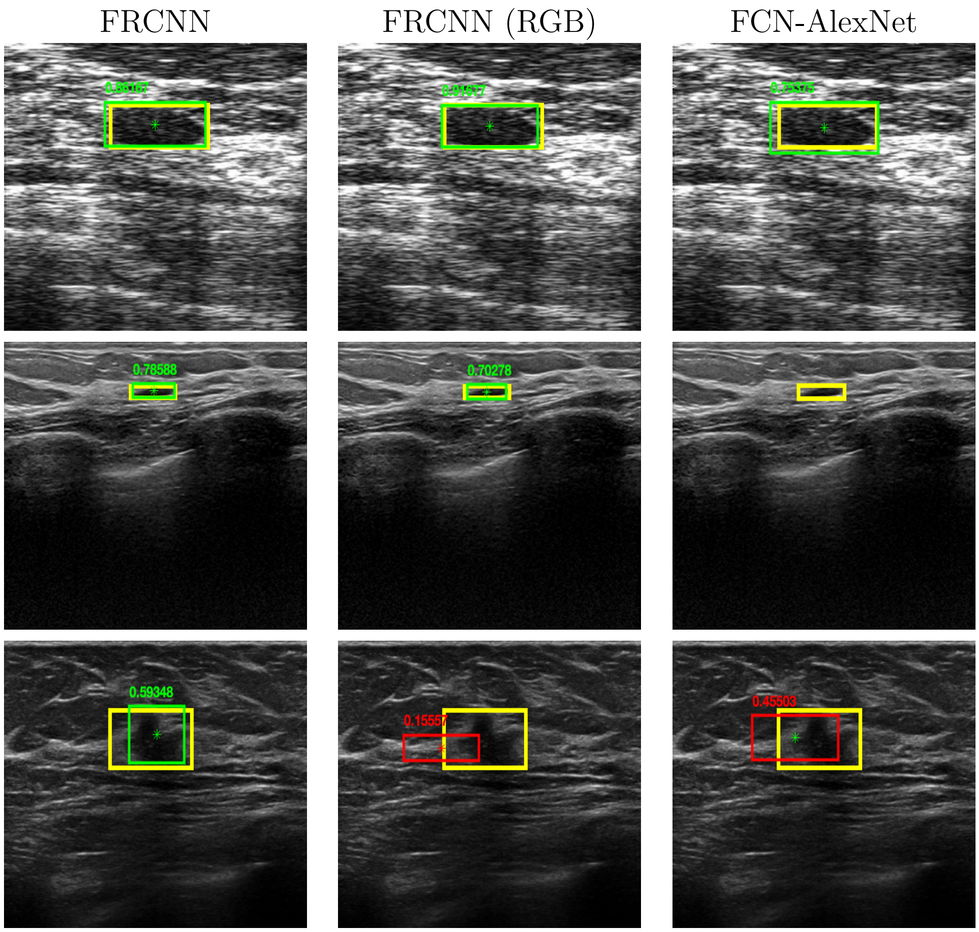


Figure 7: Examples cases from Dataset A and B to illustrate the performance of the lesion detection algorithms. The yellow rectangle indicates the ground truth, the '*' is the *detected point*, green rectangle is the *TP* and red rectangle is the *FP*. The first row (image from Dataset A) shows an easy case where all methods detected the lesion. The second row (image from Dataset B) illustrate a case where the lesion is small and only detected by FRCNN (both with and without 3-channel artificial RGB images). The third row (image from Dataset B, based on the results of composite dataset analysis) shows an image with complex shadow and all the algorithms produced different results.

- The overall performance of FRCNN was better when assessed within individual dataset (see underlined results in Table 2 and Table 3). This is an indication that it is suitable for single source datasets.

- The proposed 3-channel artificial RGB method has potential to improve the *recall* but may not be suitable for images with different resolution. In our experiment, it only performed well on Dataset A, but not Dataset B. Current results are inconclusive and required further investigation.

- The overall results of FRCNN has a higher *mean* IoU and a lower *Standard Deviation* when compared to FCN-AlexNet.

- The limitation of this paper is the comparison of FCN-AlexNet with Faster R-CNN Inception-ResNet-v2, where the differences between the two networks could be overestimated. This potential bias is due to the two very different backbones used.

5. Conclusion

This paper proposed the use of the most accurate object detection deep learning framework – Faster-RCNN with Inception-ResNet-v2 – for breast ultrasound lesion detection and localisation. It investigated the use of a 3-channel artificial RGB technique, and the applicability to transfer learning in smaller datasets. Moreover, we showed that the Faster R-CNN approach obtains the best results compared to current state of the art when evaluated on two datasets using the *detected point* measurement and overlap criterion. These were then presented in four popular metrics: *recall*, *precision*, *F1-Score* and *FPI*.

The results showed Faster-RCNN with Inception-ResNet-v2 was the most robust algorithm across two datasets and worked well on small datasets. Although FCN-AlexNet achieved good results when evaluated with *detected point*, its performances deteriorated when evaluated using the intersection over union IoU as the criterion. In addition, the new 3-channel artificial RGB technique showed improved results when evaluated on Dataset A. However, the proposed 3-channel artificial RGB technique was not suitable for either Dataset B or the composite dataset. Further areas to improve our work include:

- Investigation in using different type of image manipulation techniques will have potential in improving the use of this 3-channel artificial RGB technique.

- To overcome the limitation of this paper, the use of a different feature extraction network, such as Feature Pyramid Network (FPN ResNet-101) should be investigated to evaluate the performance of the deep learning approach.

- Increase the volume of the datasets by data collection or introducing data-augmentation techniques such as albumentation (image augmentation and composition of image augmentation).

We demonstrated the use of state-of-the-art computer vision object detection algorithm on BUS lesion localisation. This is an important step forward to improve the lesion detection of BUS. We recommended the use of IoU (equivalent to Dice Coefficient Index, which is commonly used in lesion segmentation) in lesion detection as it is more reliable when compared to the *detected point*. Our work provides an important benchmark for future research.

Acknowledgment

The authors would like to thanks Prapavesis et al. for the permission to use Dataset A.

Reference

[1] H. Cheng, J. Shan, W. Ju, Y. Guo, L. Zhang, Automated breast cancer detection and classification using ultrasound images: A survey, Pattern Recognition 43 (1) (2010) 299 – 317.

[2] [link].
URL `http://breastcancernow.org`

[3] W. Berg, L. Gutierrez, M. NessAiver, W. B. Carter, M. Bhargavan, R. Lewis, O. Ioffe, Diagnostic accuracy of mammography, clinical examination, us, and mr imaging in preoperative assessment of breast cancer, Radiology 233 (3) (2004) 830–849.

[4] K. M. Kelly, J. Dean, W. S. Comulada, S.-J. Lee, Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts, European radiology 20 (3) (2010) 734–742.

[5] A. Stavros, C. Rapp, S. Parker, Breast Ultrasound, 1st Edition, 978-0397516247, LWW, 1995.

[6] B. Huynh, K. Drukker, M. Giger, Mo-de-207b-06: Computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks, Medical Physics 43 (6) (2016) 3705–3705.

[7] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwiggelaar, A. K. Davison, R. Martí, Automated breast ultrasound lesions detection using convolutional neural networks, IEEE journal of biomedical and health informatics 22 (4) (2018) 1218–1226.

[8] M. H. Yap, E. Edirisinghe, H. Bez, Processed images in human perception: A case study in ultrasound breast imaging, European Journal of Radiology 73 (3) (2010) 682–687.

[9] W. Gómez-Flores, B. A. Ruiz-Ortega, New fully automated method for segmentation of breast lesions on ultrasound based on texture analysis, Ultrasound in medicine & biology 42 (7) (2016) 1637–1650.

[10] M. H. Yap, E. A. Edirisinghe, H. E. Bez, A novel algorithm for initial lesion detection in ultrasound breast images, Journal of Applied Clinical Medical Physics 9 (4) (2008) 181–199.

[11] K. Drukker, M. L. Giger, C. J. Vyborny, E. B. Mendelson, Computerized detection and classification of cancer on breast ultrasound, Academic Radiology 11 (5) (2004) 526–535.

[12] J. Shan, H. Cheng, Y. Wang, Completely automated segmentation approach for breast ultrasound images using multiple-domain features, Ultrasound in Medicine and Biology 38 (2) (2012) 262–275.

[13] G. Pons, R. Martí, S. Ganau, M. Sentis, J. Martí, A feasibility study of lesion detection using deformable part model in breast ultrasound images, in: Iberian Conference on Pattern Recognition and Image Analysis, Vol. 7887, 2013, pp. 269–276.

[14] M. Byra, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, M. Andre, Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion, Medical physics (2018).

[15] M. H. Yap, M. Goyal, F. M. Osman, R. Martí, E. Denton, A. Juette, R. Zwiggelaar, Breast ultrasound lesions recognition: end-to-end deep learning approaches, Journal of Medical Imaging 11007 (2019) 1.

[16] M. H. Yap, E. Edirisinghe, H. Bez, Computer aided detection and recognition of lesions in ultrasound breast images, in: Innovations in Data Methodologies and Computational Algorithms for Medical Applications, IGI Global, 2012, pp. 125–152.

[17] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning., in: AAAI, Vol. 4, 2017, p. 12.

[18] Y.-L. Huang, D.-R. Chen, Y.-K. Liu, Breast cancer diagnosis using image retrieval for different ultrasonic systems, in: Image Processing, 2004. ICIP'04. 2004 International Conference on, Vol. 5, IEEE, 2004, pp. 2957–2960.

[19] F. M. Osman, M. H. Yap, The effect of filtering algorithms for breast ultrasound lesions segmentation, Informatics in Medicine Unlocked 12 (2018) 14–20.

[20] M. A. Kupinski, M. L. Giger, Automated seeded lesion segmentation on digital mammograms, IEEE Transactions on medical imaging 17 (4) (1998) 510–517.

[21] K. Drukker, M. L. Giger, K. Horsch, M. A. Kupinski, C. J. Vyborny, E. B. Mendelson, Computerized lesion detection on breast ultrasound, Medical Physics 29 (7) (2002) 1438–1446.

[22] M. H. Yap, E. A. Edirisinghe, H. E. Bez, Object boundary detection in ultrasound images, in: The 3rd Canadian Conference on Computer and Robot Vision (CRV'06), IEEE, 2006, pp. 53–53.

[23] B. Liu, H. Cheng, J. Huang, J. Tian, X. Tang, J. Liu, Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images, Pattern Recognition 43 (1) (2010) 280–298.

[24] Y. Ikedo, D. Fukuoka, T. Hara, H. Fujita, E. Takada, T. Endo, T. Morita, Development of a fully automatic scheme for detection of

masses in whole breast ultrasound images, Medical physics 34 (11) (2007) 4378–4388.

[25] G. Pons, R. Martí, S. Ganau, M. Sentís, J. Martí, Computerized detection of breast lesions using deformable part models in ultrasound images, Ultrasound in Medicine & Biology 40 (9) (2014) 2252–2264.

[26] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[27] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, arXiv preprint arXiv:1409.0575 (2014).

[28] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, in: European conference on computer vision, Springer, 2014, pp. 740–755.

[29] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, J. Spragg, Fully convolutional networks for diabetic foot ulcer segmentation, in: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2017, pp. 618–623.

[30] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, et al., Speed/accuracy tradeoffs for modern convolutional object detectors, in: IEEE CVPR, Vol. 4, 2017.

[31] M. Goyal, N. D. Reeves, S. Rajbhandari, M. H. Yap, Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices, IEEE journal of biomedical and health informatics 23 (4) (2018) 1730–1741.

[32] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in neural information processing systems, 2015, pp. 91–99.

[33] C. L. Zitnick, P. Dollár, Edge boxes: Locating object proposals from edges, in: European conference on computer vision, Springer, 2014, pp. 391–405.

[34] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM international conference on Multimedia, ACM, 2014, pp. 675–678.

[35] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[36] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[37] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.