Article Arrival Date 13.05.2021 | Article Type Research Article | Article Published Date 20.06.2021
Doi Number: http://dx.doi.org/10.38063/ejons.428

DETECTION AND SEGMENTATION OF BREAST TUMOR LESIONS IN ULTRASOUND IMAGES WITH MASK R-CNN

ULTRASON GÖRÜNTÜLERDE GÖĞÜS TÜMÖR LEZYONLARININ MASK R-CNN İLE TESPİTİ VE BÖLÜTLENMESİ

Ahsen AYDIN BÖYÜK

Kocaeli University, Faculty of Technology, Department of Biomedical Engineering, Kocaeli, Turkey, ORCID: 0000-0002-4621-726X

Mustafa BÖYÜK

Kocaeli University, Aircraft Electrical and Electronics Department, Kocaeli, Turkey,

ORCID: 0000-0002-1196-7634

Emine DOĞRU BOLAT

Kocaeli University, Faculty of Technology, Department of Biomedical Engineering, Kocaeli, Turkey, ORCID: 0000-0002-8290-6812

ABSTRACT

Worldwide, breast cancer appears to be the type of cancer with the highest cancer mortality rates in women. As is known, the main way to reduce such death rates is through early and accurate diagnosis. In recent years, researchers have focused on convolutional neural networks-based computer vision techniques to shorten the time of diagnosis. By training neural network models with hundreds or even thousands of different breast ultrasound images, they aimed to detect the tumor area. Similarly, the main purpose of this study is to create a model for automatic detection, classification (benign-malignant) and segmentation of the lesion in ultrasound images. This model can be integrated with the PACT system of hospitals and is expected to support physicians for diagnosis in the coming years. Benign-malignant lesion differentiation was achieved by using mask regions with a deep learning model called Mask R-CNN. In addition, 4 different feature extracting backbones (ResNet50 FPN, ResNet50 C4, ResNet101 FPN, ResNet101 C4) were utilized. For Benign class, Resnet 50 C4 model achieved the highest detection in terms of AP. Resnet 101 C4 model achieved the highest performance for malignant class.

Keywords —breast cancer; deep learning; mask R-CNN; ultrasound, segmentation

ÖZET

Dünya çapında meme kanseri, kadınlarda kanserden ölüm oranlarının en yüksek olduğu kanser türü olarak görünmektedir. Bilindiği üzere bu türden ölüm oranlarını azaltmanın temel yolu erken ve doğru teşhisten geçmektedir. Son yıllarda araştırmacılar, tanı süresini kısaltmak için evrişimli sinir ağları tabanlı bilgisayarlı görme tekniklerine odaklanmışlardır. Sinir ağı modellerinin yüzlerce hatta binlerce farklı meme ultrasonu görüntüsü ile eğitilmesiyle tümörlü bölgenin tespitini hedeflemişlerdir. Bu çalışmadaki temel amaçta benzer şekilde memede oluşan iyi ya da kötü huylu tümörlerin ultrason görüntüleri kullanılarak lezyonun otomatik olarak tespiti, sınıflandırılması (iyi-kötü huylu) ve bölütlemesi için bir model oluşturmaktır. Bu model hastanelerin PACT sistemiyle entegre edilebilir ve önümüzdeki yıllarda hekimlere tanı koymaları için destek vermesi beklenmektedir. İyi-kötü huylu lezyon ayrımı maske bölgeleri kullanılarak Mask R-CNN denilen bir derin öğrenme modeli ile gerçekleştirilmiştir. Ayrıca 4 farklı özellik çıkarıcı omurga (ResNet50 FPN-ResNet50 C4-ResNet101 FPN-ResNet101 C4) kullanılmıştır. Benign sınıfı için, Resnet 50 C4 modeli

AP açısından en yüksek algılamaya ulaşmıştır. Resnet 101 C4 modeli, malign sınıf için en yüksek performansı elde etmiştir.

Anahtar Kelimeler — meme kanseri; derin öğrenme; mask R-CNN; ultrason, bölütleme

1. INTRODUCTION

All over the world, breast cancer is one of the most frequently encountered deadly cancers, especially in women [1]. It is very important to have information about the structure and mass of lesions in the breast to determine whether there is cancer in the tissue. While benign tumors which is noncancerous tend to grow slowly and do not spread throughout the body, malignant tumors that is cancerous are invasive and can invade and destroy surrounding normal tissues and spread throughout the body [2,3]. Breast Ultrasound imaging is one of the most frequently used mass diagnosis methods to detect abnormal lesions in the breast. Ultrasonic image is formed by sending ultrasonic sound waves to the tissue to be examined and reflecting this sound wave from the tissues in different periods and at different rates. Although there are methods such as mammography and magnetic resonance imaging (MRI) for imaging breast lesions, [4-7] the most preferred one is breast ultrasound (BUSI) imaging. Because it does not require intervention (non-invasive), does not use radiation-emitting agents, and obtains real-time and high-quality images. In addition, in cases where biopsy is required, ultrasonic imaging systems are very useful in real-time imaging of the tissue where the biopsy needle will be placed [5,8].

As with other biomedical imaging, the ultrasound imaging method is also affected by noise and artifacts, which reduces the detail in the target area to be detected in the image [9,10]. When the radiologist detects any tumor on the breast ultrasound image, it approximately surrounds the lesion, that is, manually tries to distinguish the area of the lesion from the background [11]. Furthermore, the expert with his/her own experience; He/she expresses the opinion that the tumor is malignant or benign by examining at many features such as the shape and structure of the lesion [12]. For all those reasons, investigating ultrasound images is a field that requires expertise, and it is quite difficult for even an expert in the field of image interpretation to obtain clear information about the structure and dimensions of the lesion due to the noise in the image [13,14]. This manual process can result in the generation of highly false negative and false positive numbers. However, with the emergence of deep learning methods recently, the development of Convolutional Neural Network (CNN) models, which can be considered quite popular, is intensely encountered in region of interest (ROI) detection in biomedical images [15] . Due to these methods, it is aimed to reduce the rate of misinterpretation or incomplete interpretation of breast lesions caused by the specialist. Therefore, great steps have been taken in early and accurate diagnosis of breast cancer. Since the process of manually extracting features from images with machine learning techniques, which are frequently used in recent years, requires expertise. However, deep learning models can automatically extract features from the images with convolutional neural networks [16-18]. In addition, this method provides very high accuracy compared to machine learning techniques.

Convolutional neural networks (CNNs), one of the deep learning networks, are often used in image classification, object detection and segmentation where a lot of images may be needed to train entire network [15].In particular, obtaining and labeling biomedical images is a very expensive process, both due to the privacy of patients and it is a field that requires expertise. Therefore, the number of biomedical images is much less than natural images. In cases where there is not enough data or images to train, using pre-trained weights instead of training the model from scratch helps to reach the result quickly. This process is called transfer leaning. Many CNN models such as VGG16 [19], DenseNet [20] Alexnet [21] are widely used to extract features from images. Not only that, but also researchers have frequently started to apply CNN-based deep learning models in different medical image processing areas such as feature extraction from the image, segmentation of the lesion, measuring lesion dimensions in the medical image and finding region of interest (ROI).

Moon et al. used different CNN architectures (DenseNet, ResNet, VGGNet) to diagnose tumors from ultrasound images [22]. There are 1687 lesion images in their own dataset. They performed the classification of malignant-benign lesions on breast ultrasound images. With their proposed CNN model, they achieved the following results: AUC, accuracy, sensitivity, specificity, precision, and F1 score; They were 0.9697, 91.10%, 85.14%, 95.77%, 94.03%, 89.36% respectively. In [23], Breast Ultrasound Images were shared publicly under the name BUSI dataset. This dataset consists of a total of 697 ultrasound images, including malignant, benign, and normal lesions. First, they developed a classification model with NasNet [24], then they increased the size of the data set using the DAGAN [25] method and retrained the model, finally they got the AUC, accuracy, sensitivity, specificity, precision, and F1 score as 0.9711 94.62%, 92.31%, 95.60%, 90%, 91.14%, respectively[26]. As a result, it has been observed that increasing the size of the dataset with different methods is very useful in improving the accuracy of the model.

In the study conducted by Chiao et al., 307 breast ultrasound images were collected from patients, and the labels corresponding to these lesions were biopsied to confirm whether they were malignant or benign [27]. Using the Mask R-CNN technique based on deep learning, they developed a model that automatically detects, classifies and segments breast lesion images. The proposed model can distinguish malignant and benign tumors with 85% accuracy, so the diagnosis can be made on breast lesions automatically.

Tanaka et al. developed a model with VGG192 and ResNe1523 architectures to diagnose malignant-benign from breast ultrasound images [28]. They used a total of 1536 breast ultrasound images with pathologically confirmed malignant-benign lesions. Their convolutional neural network showed 0.951 AUC score and 90.9 % sensitivity.

Masud et al. aimed to make breast cancer classification on ultrasound images using two different datasets [16]. The main purpose of using two different datasets is to increase the number of images. They leveraged eight different pre-trained weights which are Xception, VGG16, ResNet50, ResNet18, MobileNet, GoogleNet, Darknet19, AlexNet respectively to classify tumors. Among them, they achieved the highest accuracy rate of 92.4% with ResNet50 and 0.97 AUC with VGG16.

Researchers have carried out many studies to determine the ROI on ultrasound images using different deep learning models. Our aim in this study is to perform tumor area detection and segmentation with deep learning methods and to measure the effect of different feature extraction backbones on the performance of the Mask R-CNN model. In section 2 content of breast ultrasound image dataset and details of the ROI detection models are introduced. In Section 3, experimental results are provided, and conclusions are given in Section 4.

2. MATERIALS AND METHODS

2.1 Breast Ultrasound Image Dataset

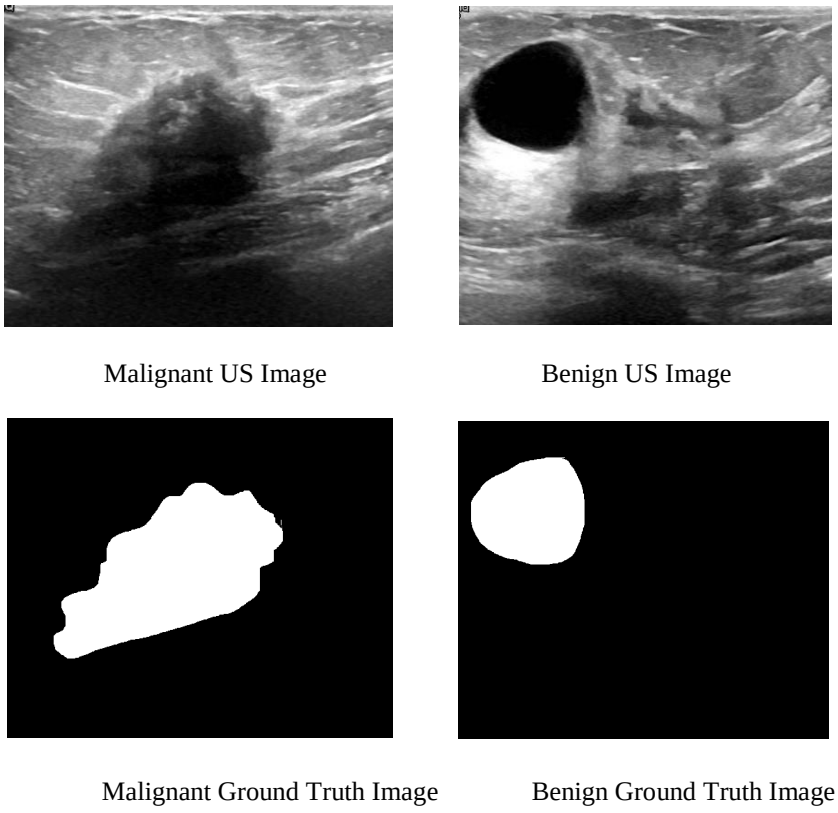
In this study, a deep learning model was developed using the breast ultrasound images that are available to the public [23]. It was stated that the data were collected in 2018 and belong to 600 women between the ages of 25-75. This dataset, in PNG format, contains 780 images with an average size of 500 x 500 pixels. In addition, it includes ground truth masks. This dataset consists of 3 classes that contain images of benign malignant and normal lesions. The distribution of dataset is given in Table 1. In this study, we only used images containing malignant and benign lesions in order to perform segmentation. For this purpose, we leveraged 697 ultrasound images in total. The total dataset is divided into two parts for 80 % training and 20 % testing. Figure 1 shows the samples in the dataset according to the classes.

A huge amount of data is needed to train deep learning models. Therefore, having enough data will affect the success of the model. In cases where the amount of data is limited, some techniques such as flipping, rotating, scaling the original image can be considered as data augmentation. There are many methods for data augmentation such as adding salt and pepper noise, flipping, rotating, rotating, scaling, zooming in, zooming out etc. It is not always necessary to apply more complex data

augmentation methods in biomedical images. Because these methods have been developed to be applied on natural images [29-31]. Therefore, we increased the amount of data by simple techniques such as random rotating, cropping, shearing and translating images at certain angles. The new augmented data obtained in this way are shown in Figure 2. Unlike the traditional data augmentation methods, there are also studies on complex data augmentation with elastic deformation in ultrasound images [32]. It has been stated that this method gives better results than traditional data augmentation methods in some datasets.

Table 1. Breast ultrasound image dataset (BUSI) [23].

Class	**Train**	**Test**	**Total**
Malignant	168	42	210
Benign	390	97	487
Normal	-	-	133



Malignant US Image Benign US Image

Malignant Ground Truth Image Benign Ground Truth Image

Figure 1. Samples of BUSI and Ground Truth Images [23].

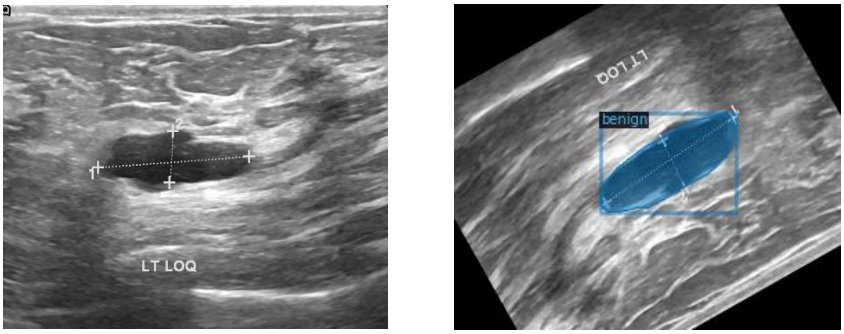


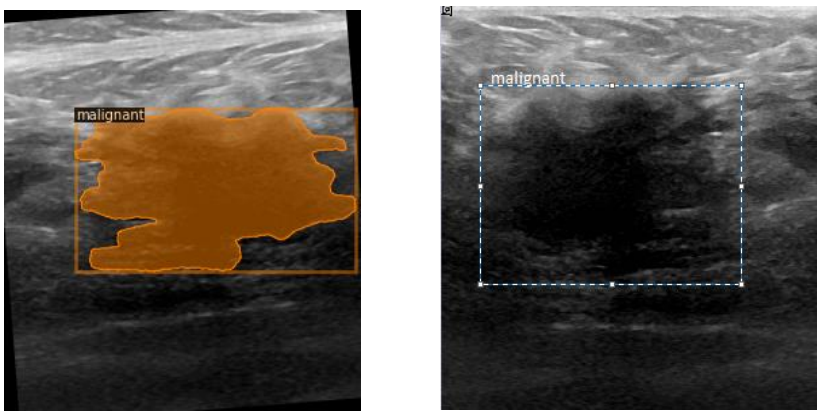
Figure 2. Original image and augmented image [23]

2.2 Contouring tumor and ground truth image

After the breast ultrasound images are collected, the procedure is to separate the lesions into different categories (malignant or benign) by observing at the structure and distribution of the lesions in these images by expert radiologists. In fact, the accuracy of the diagnosis is increased not only with the expert knowledge but also with the extra biopsy results. The dataset we used in this study also includes images that are distinguished between malignant and benign lesions and ground truth masks are given along with them.

2.3 Introduction to Deep Learning and Mask R-CNN

With the developing technology, the use of deep learning algorithms has increased considerably in the field of medicine, as in many other fields. Many processes can be automated, especially with computer vision techniques. For example, body fat measurement can be made with images taken from the camera. There are many other computer vision applications, not only object detection, but also instance segmentation, which is a computer aided technology especially in the diagnosis of cancer, is a very vital study in order to gain information about the structure and size of the lesion. If we consider the tumor diagnosis, the object detection algorithm only gives us information about the presence or absence of the object (it encapsulates the detected object in the image and does not give information about its shape). The instance segmentation also gives information about the shape of this object by creating a pixel-wised mask [33,34]. The difference between object detection and segmentation is illustrated in Figure 3.



(a) (b)

Figure 3. (a) instance segmentation (b) object detection

Segmentation can be done with many approaches (Region-Based Segmentation, Edge Detection Segmentation, Segmentation based on Clustering, etc.). However, it is more appropriate to basically divide this into two different techniques. The first of these is known as semantic segmentation and

the second as instance segmentation. Both of those can describe an object in the image and provide information about its location. The difference between these techniques becomes clear when there is more than one object in an image. For example, let us assume that there is a breast ultrasound image and there are multiple independent lesions in this image. Semantic segmentation subjects all lesions in this image to a single class and performs a segmentation between lesions and the background. To visualize segmentation, it assigns color to the pixels of the lesion to separate the lesion from the background. However, the pixels of these lesions are expressed in a single color. In other words, it actually makes segmentation as if there is a single lesion rather than more than one lesion. However, in instance segmentation, unlike semantic segmentation, each lesion assigns a class of pixels, but here the color of each lesion assigned to the same class is different from each other [35,36]. This can give us more clear information about the number of lesions in the image. If we consider this in the field of cancer diagnoses in medicine, if there are more than one lesion in an image, they will be segmented separately, and it will be very useful in examining the metastasis conditions. Mask R-CNN, developed by Facebook AI, is one of the widely used deep neural networks that aims to solve the instance segmentation problem in computer vision [37]. Although this deep neural network is an architecture suitable for simple and general applications, it provides us with the bounding box coordinates of each object in the image, label of the class and the segmentation mask of the objects.

Before explaining the Mask R-CNN model, it is necessary to understand the Faster R-CNN. Since the Mask R-CNN model is created by adding number of convolutional layers which are responsible for creating masks on the top of Faster R-CNN. Faster R-CNN consists of three neural networks which are Feature network, Region Proposal Network (RPN), and Detection Network respectively. The feature network is responsible for extracting features by using convolutional layers. Therefore, this layer can be replaced by different backbones and used to extract deep and shallow features. RPN's task is to produce bounding boxes called region of interest that contain possible objects. At the end of RPN, multiple bounding boxes are created and neglecting the boxes that are least likely to contain objects is called non-maximum suppression. After RPN, regions containing different sized CNN feature maps are suggested. Working with different sized structures is not an easy step. Therefore, this problem is solved by reducing the feature maps to the same size with the help of region of interest pooling. Detection Network is responsible for final prediction of class and bounding box coordinates, receiving inputs from both RPN and Feature Network. In Figure 4, an explanatory diagram of the Faster R-CNN is given [38].

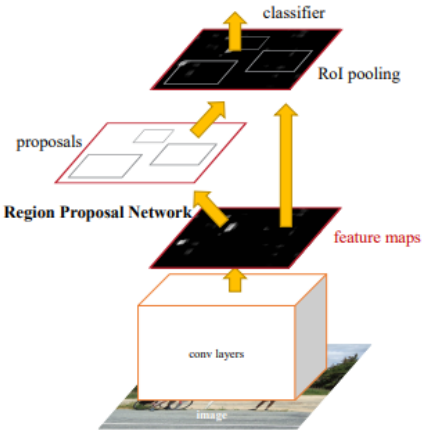


Figure 4. Faster R-CNN [39]

The basis of the Mask R-CNN technique is also the same as Faster R-CNN except for a few extensions. Mask RCNN has two stages. In stage 1, it generates proposals about regions (RPN) in which an object can exist, based on the input image. In stage 2, head part predicts the class of object, corrects the bounding box, and creates a pixel-level mask of the object according to the stage one proposal. Both stages are connected to backbone layer. The backbone is a deep neural network in the form of FPN. It includes a bottom-up pathway, a top-down pathway, and lateral connections. The

bottom-up path can usually be any convolutional neural network that serves to extract features from raw images, such as ResNet or VGG. The top-down path creates a feature pyramid map similar in size to the bottom-up path. The lateral connections combine the two corresponding pathways by convolution. Feature Pyramid Network [40] (FPN) performs better in other convolutional networks due to maintaining strong semantic properties at various resolution scales. In our study, we leveraged not only FPN backbone but also ResNet Conv4 backbone which is original baseline in Faster R-CNN to compare the detection results. A major contribution of Mask-RCNN switches the somewhat inexact ROI-Pooling process in Faster-RCNN with an operation called ROI-Align, which enables the creation of very precise instance segmentation masks. As shown in Figure 5, Mask-RCNN adds two fully convolutional neural network layers after ROI-Align layer to build the desired instance segmentations [38,41,42]. In the training phase, the loss function represents the difference between the predicted value and the ground truth. In our breast ultrasound image segmentation network, the model is trained using bounding box regression, classification, and mask prediction loss functions. The IoU value plays a significant role in measuring the success of the object detection model. Intersection over Union (IoU) is calculated with the formula given below [43,44].

IoU = Area of the intersection / Area of the union

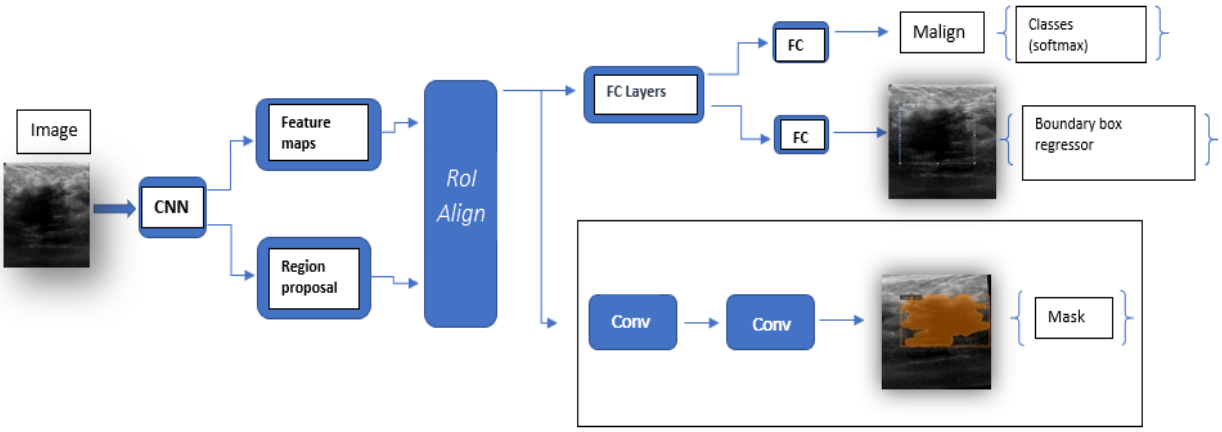


Figure 5. Mask R-CNN [46]

As a result, Faster R-CNN gives information about the class and bounding box coordinates of the objects in the image, while the Mask R-CNN architecture also gives the object mask of each object. This allows multiple objects in the image to be segmented and classified separately and to determine the individual pixels for each.

3. Experimental Results

In this study, model training and testing processes were carried out in the Google Colab environment, where it offers a free GPU to developers. With the Colab cloud system, it enables to develop deep learning applications in Python programming language using libraries such as TensorFlow, PyTorch, Keras and OpenCV. The most important advantage of Colab is that it provides high computing power independent of any computer hardware, thus helping to reach results quickly in training processes that require large amounts of processor power. The training process was measured as approximately one hour for each model. The learning rate was initially determined as 0.0025, after 1000 steps it was reduced by 5% in every 100 steps, and the total training time was determined as 1500 steps. Training of the deep learning model was carried out with Detectron2-Pytorch Framework [45]. Detectron2 Framework contains a wide variety of object detection models, but also keeps the tagged data in Json

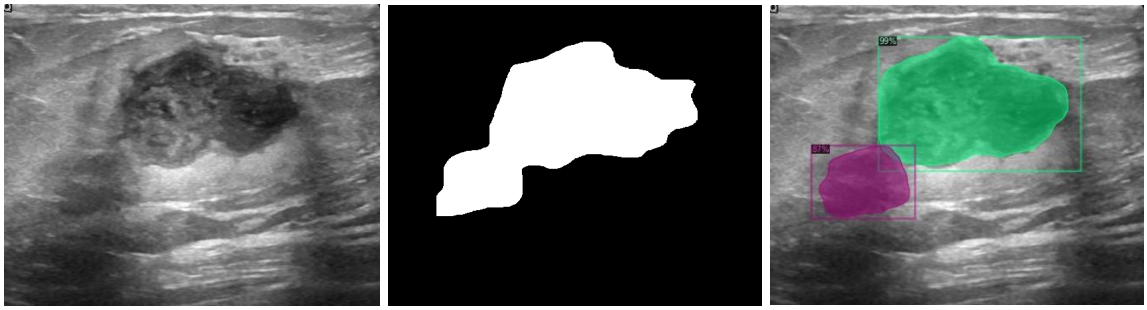
format. Therefore, the binary masks in the dataset were first converted to polygon format and then stored in json format. Batch number 6 has been selected. The COCO evaluation metric was used to evaluate the success of the model and the results were obtained in terms of AP (average precision).
The symbols APb@0.5 and APb@0.75 represent bounding boxes or predicted mask results for different IoU threshold values, while the letters m and b represent masks and bounding boxes. According to the COCO evaluation metric, AP value is calculated by increasing the IoU threshold from 0.5 to 0.95 by the amount of 0.05 and taking the average of the calculated value for each threshold.
In this study, ResNet-50 FPN, ResNet-50 C4, ResNet-101 FPN and ResNet-101 C4 backbones are used as a feature extractor in Mask R-CNN. The expressions specified here with C4 and FPN are pre-trained backbone weights which are used in training and testing phase. For Benign class, Resnet 50 C4 model achieved the highest detection in terms of AP while, Resnet 101 C4 model achieved the highest performance for malignant class. All the results for the detection and instance segmentation obtained from the BUSI dataset for each backbone are given in Table 2. The results of the instance segmentation indicated with AP-m, and those indicated with AP-b are the results of object detection. Here 'm' stands for mask, 'b' stands for bounding box.

Table 2. Experimental results

Dataset	Backbone	AP-m	AP-m@0.5	AP-m@0.75	AP-b	AP-b@0.5	AP-b@0.75
Benign	R50 FPN	57.285	78.822	69.520	56.911	79.102	67.245
Malignant	R50 FPN	40.220	74.299	37.363	40.099	75.310	37.307
Benign	R50 C4	58.414	80.351	68.267	59.195	80.351	69.600
Malignant	R50 C4	37.524	71.891	32.061	39.321	71.891	25.341
Benign	R101 C4	57.181	78.69	67.752	56.982	78.780	66.153
Malignant	R101 C4	45.407	80.14	51.08	44.2	80.460	41.055
Benign	R101 FPN	57.497	80.508	68.160	59.181	80.667	67.303
Malignant	R101 FPN	39.021	74.102	33.038	39.947	74.793	40.601

The raw ultrasound image and mask of a malignant lesion and segmentation results are illustrated in Figure 6. As seen in c2, although there was only one malignant lesion in the original image, the trained model predicted it as 2 different malignant lesions. The fact that the CNN failed to establish the connection between body and tail of the tumor. However, in Figure 6-c1, network obtained more correct segmentation result.

a1 b1 c1



a2 b2 c2

Figure 6. a1-a2) Original images, b1-b2) Original mask, c1-c2) Prediction of malignant lesion

4. Discussion and Conclusion

Breast ultrasound imaging is frequently used in the diagnosis of breast cancer, which is seen at a high rate, especially in women, worldwide. Therefore, early and rapid diagnosis is of vital importance for the patient's life. Deep neural network was utilized for automatic detection, classification and segmentation of the lesion area. Since the size of dataset used in the study is small, transfer learning method was employed. The Mask R-CNN, which was developed by the Facebook artificial intelligence team was used as a head of neural network, and the ResNet50 FPN-C4 and ResNet101 FPN-C4 backbones are employed as a feature extractor. Moreover, for the benign class, the resnet50-C4 backbone achieves the highest result as 58.4 Ap, for the malignant class this value was obtained with the Resnet 101c4 backbone is 45.407 AP. It has been observed that the segmentation and object detection model give higher results in benign images. This difference may be caused by many reasons such as the number of benign lesions being higher than malignant ones. Moreover, if the IOU threshold value is determined as 0.5, the segmentation results in both categories are quite high. In our future study, we plan to apply different data augmentation methods such as GAN (Generative Adversarial Network), elastic deformation, to increase the number of samples in training set and compare them with traditional data augmentation.

5. References

[1] S.-H. Shieh, V. C.-R. Hsieh, S.-H. Liu, C.-R. Chien, C.-C. Lin, and T.-N. Wu, "Delayed time from first medical visit to diagnosis for breast cancer patients in Taiwan.," *J. Formos. Med. Assoc.*, vol. 113, no. 10, pp. 696–703, Oct. 2014, doi: 10.1016/j.jfma.2012.12.003.

[2] M. Wei *et al.*, "A Benign and Malignant Breast Tumor Classification Method via Efficiently Combining Texture and Morphological Features on Ultrasound Images," *Comput. Math. Methods Med.*, vol. 2020, 2020, doi: 10.1155/2020/5894010.

[3] M. Byra *et al.*, "Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion," *Med. Phys.*, vol. 46, no. 2, pp. 746–755, 2019, doi: 10.1002/mp.13361.

[4] Y. Ouyang, P. H. Tsui, S. Wu, W. Wu, and Z. Zhou, "Classification of benign and malignant breast tumors using h-scan ultrasound imaging," *Diagnostics*, vol. 9, no. 4, pp. 1–11, 2019, doi: 10.3390/diagnostics9040182.

[5] H. D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, "Automated breast cancer detection and classification using ultrasound images: A survey," *Pattern Recognit.*, vol. 43, no. 1, pp. 299–317, 2010, doi: 10.1016/j.patcog.2009.05.012.

[6] H. Zhi, B. Ou, and B. Luo, "Comparison of Ultrasound Elastography, Mammography, and Sonography in the Diagnosis of Solid Breast Lesions," *Ultrasound*, pp. 807–815, 2010.

[7] K. Drukker, M. L. Giger, K. Horsch, M. A. Kupinski, C. J. Vyborny, and E. B. Mendelson, "Computerized lesion detection on breast ultrasound," *Med. Phys.*, vol. 29, no. 7, pp. 1438–1446, 2002, doi: 10.1118/1.1485995.

[8] F. Gharekhanloo, M. M. Haseli, and S. Torabian, "Value of ultrasound in the detection of benign and malignant breast diseases: A diagnostic accuracy study," *Oman Med. J.*, vol. 33, no. 5, pp. 380–386, 2018, doi: 10.5001/omj.2018.71.

[9] S. H. Contreras Ortiz, T. Chiu, and M. D. Fox, "Ultrasound image enhancement: A review," *Biomed. Signal Process. Control*, vol. 7, no. 5, pp. 419–428, 2012, doi: 10.1016/j.bspc.2012.02.002.

[10] J. Shan, H. D. Cheng, and Y. Wang, "Completely automated segmentation approach for breast ultrasound images using multiple-domain features," *Ultrasound Med. Biol.*, vol. 38, no. 2, pp. 262–275, 2012, doi: 10.1016/j.ultrasmedbio.2011.10.022.

[11] M. H. Yap, E. A. Edirisinghe, and H. E. Bez, "A novel algorithm for initial lesion detection in ultrasound breast images," *J. Appl. Clin. Med. Phys.*, vol. 9, no. 4, pp. 181–199, 2008, doi: 10.1120/jacmp.v9i4.2741.

[12] G. Pons, R. Martí, S. Ganau, M. Sentís, and J. Martí, "Feasibility study of lesion detection using deformable part models in breast ultrasound images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7887 LNCS, pp. 269–276, 2013, doi: 10.1007/978-3-642-38628-2_32.

[13] W. C. Shia, L. S. Lin, and D. R. Chen, "Classification of malignant tumours in breast ultrasound using unsupervised machine learning approaches," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-81008-x.

[14] K. H. Hwang *et al.*, "Computer aided diagnosis (CAD) of breast mass on ultrasonography and scintimammography," *Proc. 7th Int. Work. Enterp. Netw. Comput. Healthc. Ind. Heal. 2005*, pp. 187–190, 2005, doi: 10.1109/HEALTH.2005.1500435.

[15] J. Xing *et al.*, "Using BI-RADS Stratifications as Auxiliary Information for Breast Masses Classification in Ultrasound Images," *IEEE J. Biomed. Heal. Informatics*, vol. XX, no. XX, pp. 1–1, 2020, doi: 10.1109/jbhi.2020.3034804.

[16] M. Masud, A. E. Eldin Rashed, and M. S. Hossain, "Convolutional neural network-based models for diagnosis of breast cancer," *Neural Comput. Appl.*, vol. 5, 2020, doi: 10.1007/s00521-020-05394-5.

[17] A. Ghoneim, G. Muhammad, and M. S. Hossain, "Cervical cancer classification using convolutional neural networks and extreme learning machines," *Futur. Gener. Comput. Syst.*, vol. 102, pp. 643–649, 2020, doi: 10.1016/j.future.2019.09.015.

[18] F. Hu, G. S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, 2015, doi: 10.3390/rs71114680.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2015, Accessed: Apr. 29, 2021. [Online]. Available: http://www.robots.ox.ac.uk/.

[20] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2261–2269, Aug. 2016, Accessed: Apr. 29, 2021. [Online]. Available: http://arxiv.org/abs/1608.06993.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks." Accessed: Apr. 29, 2021. [Online]. Available: http://code.google.com/p/cuda-convnet/.

[22] W. K. Moon, Y. W. Lee, H. H. Ke, S. H. Lee, C. S. Huang, and R. F. Chang, "Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 190, 2020, doi: 10.1016/j.cmpb.2020.105361.

[23] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Br.*, vol. 28, p. 104863, 2020, doi: 10.1016/j.dib.2019.104863.

[24] B. Zoph, G. Brain, V. Vasudevan, J. Shlens, and Q. V Le Google Brain, "Learning Transferable Architectures for Scalable Image Recognition."

[25] A. Antoniou, A. Storkey, and H. Edwards, "Data Augmentation Generative Adversarial Networks," *arXiv*, Nov. 2017, Accessed: Apr. 29, 2021. [Online]. Available: http://arxiv.org/abs/1711.04340.

[26] W. Al-Dhabyani, A. Fahmy, M. Gomaa, and H. Khaled, "Deep learning approaches for data augmentation and classification of breast masses using ultrasound images," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 618–627, 2019, doi: 10.14569/ijacsa.2019.0100579.

[27] J. Y. Chiao, K. Y. Chen, Ken Ying-Kai Liao, P. H. Hsieh, G. Zhang, and T. C. Huang, "Detection and classification the breast tumors using mask R-CNN on sonograms," *Med. (United States)*, vol. 98, no. 19, pp. 1–5, 2019, doi: 10.1097/MD.0000000000015200.

[28] H. Tanaka, S.-W. Chiu, T. Watanabe, S. Kaoku, and T. Yamaguchi, "Computer-aided diagnosis system for breast ultrasound images using deep learning," *Ultrasound Med. Biol.*, vol. 45, p. S4, 2019, doi: 10.1016/j.ultrasmedbio.2019.07.426.

[29] N. Aghnia Farda, J. Y. Lai, J. C. Wang, P. Y. Lee, J. W. Liu, and I. H. Hsieh, "Sanders classification of calcaneal fractures in CT images with deep learning and differential data augmentation techniques," *Injury*, vol. 52, no. 3, pp. 616–624, 2021, doi: 10.1016/j.injury.2020.09.010.

[30] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *J. Comput. Sci.*, vol. 30, pp. 174–182, 2019, doi: 10.1016/j.jocs.2018.12.003.

[31] T. Nemoto *et al.*, "The Effects Of Sample Size And Data Augmentation On The Efficacy Of Semantic Segmentation For Prostate Cancer Using Deep Learning: A Report Of More Than 500 Cases," *Int. J. Radiat. Oncol.*, vol. 108, no. 3, pp. e767–e768, 2020, doi: 10.1016/j.ijrobp.2020.07.209.

[32] E. Castro, J. S. Cardoso, and J. C. Pereira, "Elastic deformations for data augmentation in breast cancer mass detection," *2018 IEEE EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2018*, vol. 2018-Janua, no. March, pp. 230–234, 2018, doi: 10.1109/BHI.2018.8333411.

[33] M. Böyük, R. Duvar, and O. Urhan, "Deep Learning Based Vehicle Detection with Images Taken from Unmanned Air Vehicle," pp. 2–5, 2020.

[34] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[35] S. Hao, Y. Zhou, and Y. Guo, "A Brief Survey on Semantic Segmentation with Deep Learning," *Neurocomputing*, vol. 406, pp. 302–321, 2020, doi: 10.1016/j.neucom.2019.11.118.

[36] B. R. Kang, H. Lee, K. Park, H. Ryu, and H. Y. Kim, "BshapeNet: Object detection and instance segmentation with bounding shape masks," *Pattern Recognit. Lett.*, vol. 131, pp. 449–455, 2020, doi: 10.1016/j.patrec.2020.01.024.

[37] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, "Instance segmentation of apple flowers using the improved mask R–CNN model," *Biosyst. Eng.*, vol. 193, pp. 264–278, 2020, doi: 10.1016/j.biosystemseng.2020.03.008.

[38] R. Anantharaman, M. Velazquez, and Y. Lee, "Utilizing Mask R-CNN for Detection and Segmentation of Oral Diseases," *Proc. - 2018 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2018*, pp. 2197–2204, 2019, doi: 10.1109/BIBM.2018.8621112.

[39] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[40] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," Dec. 2016, Accessed: Apr. 29, 2021. [Online]. Available: http://arxiv.org/abs/1612.03144.

[41] W. Jia, Y. Tian, R. Luo, Z. Zhang, J. Lian, and Y. Zheng, "Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot," *Comput. Electron. Agric.*, vol. 172, no. March, p. 105380, 2020, doi:

10.1016/j.compag.2020.105380.
[42] Y. Qiao, M. Truman, and S. Sukkarieh, "Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming," *Comput. Electron. Agric.*, vol. 165, no. March, p. 104958, 2019, doi: 10.1016/j.compag.2019.104958.
[43] F. Hou, W. Lei, S. Li, J. Xi, M. Xu, and J. Luo, "Improved Mask R-CNN with distance guided intersection over union for GPR signature detection and segmentation," *Autom. Constr.*, vol. 121, no. October 2020, p. 103414, 2021, doi: 10.1016/j.autcon.2020.103414.
[44] M. Bizjak, P. Peer, and Ž. Emeršic, "Mask R-CNN for ear detection," *2019 42nd Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2019 - Proc.*, pp. 1624–1628, 2019, doi: 10.23919/MIPRO.2019.8756760.
[45] "Detectron2: A PyTorch-based modular object detection library." https://ai.facebook.com/blog/-detectron2-a-pytorch-based-modular-object-detection-library-/ (accessed Apr. 29, 2021).
[46] S. Gonzalez, C. Arellano, and J. E. Tapia, "Deepblueberry: Quantification of Blueberries in the Wild Using Instance Segmentation," *IEEE Access*, vol. 7, pp. 105776–105788, 2019.
"Detectron2: A PyTorch-based modular object detection library." .