Residual Deep Learning System for Mass Segmentation and Classification in Mammography

Conference Paper · September 2019

DOI: 10.1145/3307339.3342157

5 authors, including:

Dina Abdelhafiz
University of Connecticut

Clifford K. Yang
University of Connecticut

Residual Deep Learning System for Mass Segmentation and Classification in Mammography

Dina Abdelhafiz*†
Department of Computer Science and Engineering, University of Connecticut,
Storrs, CT, USA
dina.abdelhafiz@uconn.edu

Sheida Nabavi†
Department of Computer Science and Engineering, University of Connecticut,
Storrs, CT, USA
sheida.nabavi@uconn.edu

Reda Ammar
Department of Computer Science and Engineering, University of Connecticut,
Storrs, CT, USA
reda.ammar@uconn.edu

Clifford Yang
Department of Diagnostic Imaging, University of Connecticut Health Center,
Farmington, CT, USA
cyang@uchc.edu

Jinbo Bi
Department of Computer Science and Engineering, University of Connecticut,
Storrs, CT, USA
jinbo.bi@uconn.edu

ABSTRACT

Automatic extraction of breast mass in mammogram (MG) images is a challenging task due to the varying sizes, shapes, and textures of masses. Moreover, the density of MGs makes mass detection very challenging since masses can be hidden in dense MGs. In this paper, we propose a residual deep learning (DL) system for mass segmentation and classification in mammography. The overall proposed system consists of two cascaded parts: 1) a residual attention U-Net model (RU-Net) to precisely segment mass lesions in MG images, followed by 2) a ResNet classifier to classify the detected binary segmented lesions into benign or malignant. The proposed semantic based CNN model, RU-Net, has the basic architecture of the U-Net model, which extracts contextual information combining low-level feature with high-level ones. We have modified the U-Net structure by adding residual attention modules in order to preserve the spatial and context information, help the network have deeper architecture, and handles the gradient vanishing problem. We compared the performance of the proposed RU-Net model with those of state-of-the-art two semantic segmentation models, and two object detectors using public databases. We also examined the effect of the breast density on the accuracy of localizing and segmenting the breast masses. Our proposed model shows superior performance compared to the other DL methods in detecting and segmenting masses, especially for heterogeneously dense and dense MG images, in terms of intersection over union (IOU) and the Dice index coefficient (DI). Moreover, our results show that the cascaded ResNet model, trained using binary-scale images, classify the masses to benign or malignant with higher accuracy compared to the ResNet model that is trained on gray-scale images.

*Dina Abdelhafiz has 2^{nd} affiliation in the Informatics Research Institute (IRI), City of Scientific Research and Technological Applications (SRTA-City), Egypt.
†Corresponding author.

ACM-BCB '19, September 7–10, 2019, Niagara Falls, NY, USA

ACM ISBN 978-1-4503-6666-3/19/09…$15.00
https://doi.org/10.1145/3307339.3342157

CCS CONCEPTS

• **Theory of computation** → **Machine learning theory**; • **Applied computing** → *Imaging*; *Bioinformatics.*

KEYWORDS

Mammograms (MGs); convolutional neural networks (CNN); breast cancer; transfer learning (TL); deep learning (DL); computer aided detection (CAD); semantic segmentation; classification; ground truth maps (GTMs); yolo; faster r-cnn; u-net; res-net

ACM Reference Format:
Dina Abdelhafiz, Sheida Nabavi, Reda Ammar, Clifford Yang, and Jinbo Bi. 2019. Residual Deep Learning System for Mass Segmentation and Classification in Mammography. In *10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB '19), September 7–10, 2019, Niagara Falls, NY, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3307339.3342157

1 INTRODUCTION

Mammograms (MGs) have contributed significantly to the reduction of the breast cancer mortality rate through early detection of cancer. Recent advances in computational technologies, and significant progress in deep learning (DL) [20, 27] and image processing techniques [24] have opened up unprecedented opportunities to develop models for providing an objective view to radiologists with higher accuracy [1, 4, 39]. With advances in detection and localization methods in DL techniques for medical imaging [21], few studies have proposed DL models to localize mass lesions in MG images [3, 5, 11]. Studies in [1, 4, 11] show that convolution neural networks (CNNs) achieve higher detection accuracy in locating masses in MGs compared to traditional Computer-Aided Detection (CAD) models. Various approaches have been proposed to further improve the accuracy of deep CNNs in detecting and localizing breast abnormalities [1, 4, 39]. Further, techniques such as stochastic depth [18], batch normalization (BN) [19], transfer learning (TL) [25], data augmentation (Aug.) [20], and dropout [33] have been used in various researches for avoiding network overfitting and regularization purposes. Despite the recent advances in the structure of DL models, detection of masses in MG images has remained a challenging problem due to the following reasons: 1) existence of some masses in the pectoral muscle area, 2) hidden masses under the dense breast tissues, and 3) varying sizes, shapes, and texture of masses [1, 4, 39].

In this study, we propose a residual DL system for mass segmentation and classification in mammography. The overall proposed system consists of two cascaded parts: 1) a residual attention U-Net model (RU-Net) to precisely segment mass lesions in MG images, followed by 2) a cascaded ResNet [16] classifier to classify the detected binary segmented lesions into benign or malignant. The proposed semantic based CNN model, RU-Net, has the basic architecture of the U-Net model [32], which extracts contextual information combining low-level feature with high-level ones. In our previous work, we have shown that the basic U-Net model can be used for more precise and efficient mass segmentation in MG images [3]. To further improve the performance of the basic U-Net model for mass segmentation, we have modified its structure by adding residual attention modules. These modules generate attention-aware features that change adaptively as the network goes deep in layers. The residual modules [16] resolve the problem of vanishing gradients using identity skip-connections thus facilitating the training of the proposed model. The proposed RU-Net use long and short skip connections to produce precise and detailed segmentation maps. Besides adding the stacked residual attention modules, we used augmented data-set in the training process to improve the accuracy of the RU-Net model.

To evaluate the performance of the proposed model, we compared the performance of the proposed RU-Net model with those of the basic U-Net model [32] and the vanilla U-Net model [3]. The performance of the models is evaluated in terms of dice index coefficient (DI), intersection over union (IOU) and inference test time. We also constructed bounding boxes (BBs) surrounding the segmented lesions to compare the output BBs produced by the proposed RU-Net with the BBs produced by the state-of-the-art Faster R-CNN and Yolo object detectors proposed in [6, 31]. We used publicly available data-sets, CBIS-DDSM [17], and BCDR-D01 [22] to train the proposed RU-Net model and the models we used for performance comparison proposed in [3, 6, 31, 32], and we tested the models using the INbreast data-set [10]. These data-sets include mass lesions of different sizes, shapes, and margins. We also examined the effect of the breast density on the accuracy of localizing and segmenting the breast masses. We tested the performance of our proposed RU-Net and the models in [3, 6, 31, 32] separately on each breast density category based on the BI-RADS code: fatty, scattered, heterogeneously dense and dense breasts, in terms of the Dice index coefficient (DI) and the intersection over index (IOU). RU-Net is implemented in Matlab R2018b and is available at: https://github.com/NabaviLab/RU-Net.

2 RELATED WORK

Despite the huge success of DL methods in classifying MG images into normal, benign and malignant [1, 4, 39], the use of CNNs for segmentation and localization of lesions in MG images are not thoroughly investigated. The region-based CNN (R-CNN) models and its faster variants, Fast R-CNN, Faster R-CNN, and Mask R-CNN [14, 15, 30] for object detection have recently been used for mass localization tasks in mammography [29, 31, 34]. Yolo (You Only Look Once) that is an effective and efficient object detection DL model has been used for mass localization [5–7].

Anchor boxes are widely adopted in state-of-the-art DL object detection models (e.g. Faster R-CNN and Yolo V2). The main drawback of these models is that if anchor boxes are not chosen correctly, the model will struggle in detecting small or irregular objects. In the Yolo model, the anchor shapes are obtained by k-means clustering on the sizes of the ground truth BBs. In Faster R-CNN, the anchor shapes are of 3 scales and of 3 aspect ratios, yielding 9 different anchors at each output sliding window position. The aspect ratios used in the case of detecting general objects such as pedestrian, car, text is different from the aspect ratios used in detecting lesions in MGs, or in medical images in general. In these models, the anchor shape has to be manually modified according to the ground-truth data-set to improve the detection accuracy and to detect small mass lesions. The Faster R-CNN [30] and Yolo [28] models need to predefine the anchor box's shapes and to fix its size during training, which is sub-optimal since it ignores the augmented data distribution in training. Inappropriate anchor boxes could degrade the performance of the detector in terms of accuracy [37] and detecting small lesions.

Recently, the fully convolutional network (FCN) and its variant improved models such as U-Net [32] and SegNet [8], have yielded outstanding results for semantic segmentation of biomedical images and a promising results for segmenting lesions in MG images [3, 38]. In these studies [5, 7], in order to enhance the detection performance of the used Yolo model in term of precision in detecting masses, the authors first, used a cascaded semantic CNN to segment the detected masses. Then, they used another cascaded CNN trained on gray-scale MG images to classify the segmented masses as either benign or malignant [5, 7]. These cascaded models increase the computation cost, however, provide better classification results.

Several studies [11, 36] have proposed a patch-based CNN to detect lesions. The drawback of the patch-based approaches is that the input patches came from non-overlapping areas, which makes it difficult to precisely localize masses. Moreover, the size of the input patches is very small that produces difficulty in differentiating normal tissues from abnormal ones after detection.

In this study, we propose a residual DL system to segment and classify lesions in MG images without the need of any cascaded detectors. To date, only a few attempts based on DL have been presented for semantic segmentation of mass lesions in mammography [3, 38].

3 MATERIAL AND METHODS

3.1 Data-sets

We trained our proposed model on two data-sets, DDSM [17] and BCDR-01 [22], and tested it on the INbreast data-set [10]. CBIS-DDSM is a digitized screen-film mammography (SFM) data-set that is a subset of the DDSM data-set [17] with updated lesion segmentation, and verified pathology. We used 2,734 images from the CBIS-DDSM data-set that have mass lesions. BCDR-D01 [22] is a screen film mammography (SFM) repository of 135 MG images. INbreast [10] is public data-set for MGs which comprises fully field digital mammography (FFDM) MG images. It has 107 MGs that are annotated for masses. All images containing masses have associated pixel-level boundary of the mass lesions annotated by experienced radiologists. To have ground truth for evaluating object detection methods, we generated ground truth BBs for the masses based on minimum and maximum points values of x and y coordinates of the mass' contours, which indicate the locations of masses. Each MG image has been annotated based on their density derived from the American College of Radiology's (ACR) Breast Imaging Reporting and Data System (BI-RADS) [9]. For each MG image, its density in ACR standard scale is given as one of these categories: class A: fatty, class B: scattered, class C: heterogeneous dense, and class D: dense [1, 4, 9]. We grouped the BI-RADS multi-class assessment into benign and malignant classes. In this study, we categorized 1,133 MG images with BI-RADS $\in \{2, 3\}$ as benign, and 1,843 MG images with BI-RADS $\in \{4, 5, 6\}$ as malignant. Distribution of density for each BI-RADS class is presented in Table 1. In total, we used 2,976 MG images to conduct our experiments.

Table 1: Distribution of breast density in each BI-RADS class in the publicly available data-sets used in our study.

Purpose	Data-set	BI-RADS	Class				Total
			A	B	C	D	All
Training	**DDSM**	Benign	154	412	228	224	1,018
Training	**BCDR-D01**	Benign	34	16	23	7	80
Test	**INbreast**	Benign	12	4	13	6	35
Training	**DDSM**	Malignant	219	693	513	291	1,716
Training	**BCDR-D01**	Malignant	16	15	22	2	55
Test	**INbreast**	Malignant	30	32	8	2	72

3.2 Data Pre-processing

We first detect the breast boundary for removing a big portion of the black background [13] from the training images. After that, we employ the contrast limited adaptive histogram equalization (CLAHE) [26] to enhance the contrast of the MG images. In previous works [1–4], we have shown that the CLAHE filter performs better compared to other commonly used filters for CNN based analysis for MGs. We generated ground truth maps (GTMs) for the masses using the associated pixel-level boundary of the mass lesions given by the data-sets. All pixels in the GTM are labeled as belonging to the background (0) or breast lesion (255) classes. All full MGs images and it's corresponding GTMs are re-sized to 640×640. To deal with the small training data-set and avoiding overfitting our model, we applied data augmentation to the training MG images and it's corresponding GTMs by image rotation by (-45, 45) degrees, translation up and down by (-10%, 10%), scaling in and out by 0.2, and left-right flips.

3.3 Proposed RU-Net

U-Net is a popular end-to-end encoder-decoder network for semantic segmentation that is originally invented for bio-medical image segmentation [32]. U-Net consists of a contracting path to capture features and an asymmetric expanding path that enables precise localization and segmentation of pixels. This architecture has a U-shaped skipping structure that connects the high-resolution features from the contracting path to the up-sampled outputs of expanding path. Inspired by the residual attention mechanism proposed in [35], we built the proposed RU-Net model by stacking residual attention modules to the basic U-Net architecture. We use the residual blocks with the identity connections instead of the regular convolution layers in the U-Net architecture in order to preserve the spatial and context information, help the network have deeper architecture, and handles the gradient vanishing problem. The residual blocks directly propagate features from its early convolution to its late convolution and improve the performance of the proposed model consequently. In order to address the problem of detecting small lesions, the proposed RU-Net model uses residual attention blocks to increase the resolution for better pixel-level prediction (Table 2). The residual attention module consists of a soft mask branch and trunk branch [35] (Fig. 1). The attention residual mechanism can keep the flow of original feature information through the trunk branch using the identity mapping and construct attention to mass lesions features using the soft mask branch. Each trunk branch is connected to its own soft mask branch (Fig. 1). The proposed RU-Net network consists of multiple levels, and in each level, the network capture features with different resolutions. As shown in Fig. 2a, the encoder in the soft mask branch at each level consists of a cascade of a down-sampling layer, a convolution (Conv.) block, a residual block, another convolution block,

Table 2: Architecture of the proposed RU-Net. The symbol ↓ means that this level in the encoder path consists of a convolution block, a residual block, a convolution block, and a down-sampling layer. The symbol ↑ means that this level in the decoder path consists of a convolution block, a residual block, a convolution block, and a down-sampling layer. The symbol □ means that this level consists of a bridge of convolution block, a residual block, and another convolution block.

Layer name	Path	Layers inside	Output resolution	Output width
Input	Encoder	↓	640×640	1
Level 1	Encoder	↓	640×640	64
Level 2	Encoder	↓	320×320	128
Level 3	Encoder	↓	160×160	256
Level 4	Encoder	↓	80×80	512
Bridge		□	40×40	1024
Level 4	Decoder	↑	80×80	512
Level 3	Decoder	↑	160×160	256
Level 2	Decoder	↑	320×320	128
Level 1	Decoder	↑	640×640	64
Classifier	Decoder	↑	640×640	1

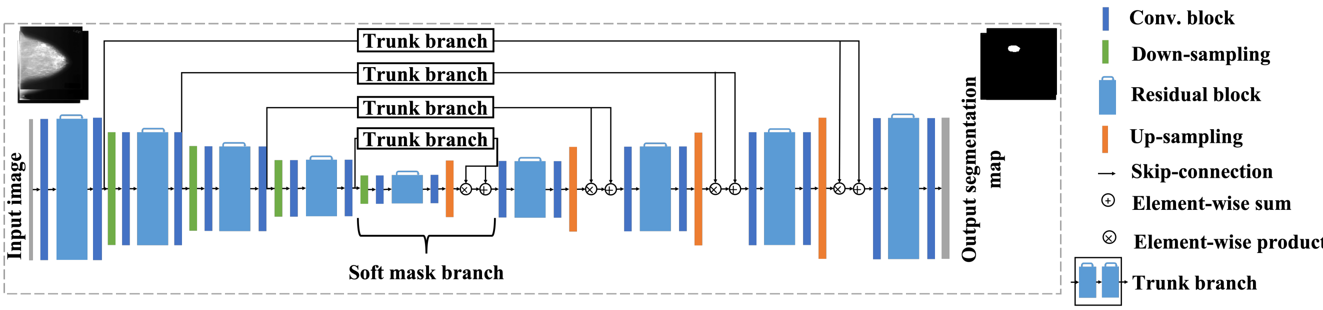
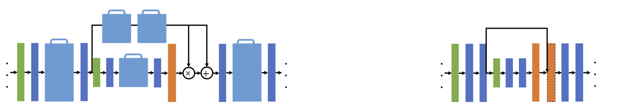


Figure 1: The architecture of the proposed RU-Net. The nested long residual skip connections connect the encoder and decoder paths at the same level, while each intermediate residual block contains a short residual skip connection within the same path to increase the depth of the proposed RU-Net model.



(a) Proposed RU-Net. **(b) Vanilla U-Net.**

Figure 2: Difference between the proposed RU-Net (a) and the vanilla U-Net (b). The proposed RU-Net model is a fully residual model that has long and short skip connections.

(a) A residual block. **(b) A convolution block.**

Figure 3: Difference between a residual block (a) and convolution block (b).

and a skip connection of residual blocks that is connected to the corresponding level in the decoder path.

The attention module keeps useful information by applying element-wise product between feature coming from the truck branch and the output of the soft mask branch. However, repeated element-wise product across layers will lead to degradation of both useful and useless information. To avoid this degradation in information across layers, an element-wise sum is then performed between the output of the element-wise product and output from the residual blocks in the trunk branch. This element-wise summation relieves the feature attenuation happened during the element-wise product process by using long connections (identical mapping), which enhances the feature contrast and improve the discrimination of the features. The output from the element-wise summation is then forwarded to the decoder path, which at each level consists of a cascade of a convolution block, a residual block, another convolution layer, and finally an up-sampling deconvolution layer. The final feature output of every residual block is the element-wise summations of the output of three cascaded convolution blocks with the short identity map (Fig. 3a). Each convolution block consists of a BN layer and an activation ReLU layer and a regular convolution layer (Fig. 3b). The down-sampling (max-pooling) layers exist between the levels in the encoding path to perform down-sampling in the feature maps. The deconvolution layers exist between levels in the decoding path to up-sample the input feature maps from the decoder level and then concatenate them using a pixel-wise addition with the feature maps coming from the encoding path by the long skip connections.

Besides the long skip connections used between each level in the encoder-decoder path, short connections are used in the residual blocks for a direct connection between layers in the same levels. Using short and long connections help the flow of information within and across levels in the RU-Net architecture to generate richer information hierarchy (Fig. 1). The trunk branch in each level use its long skip connection as input to a cascade of two residual blocks (Fig. 1). The output of the truck branch is then element-wise summed with the up-sampling feature maps from the corresponding level in the decoder path. The final segmented binary map is obtained by passing the result through a pixel-wise Sofmax classifier after the last convolution layer.

The original U-Net model uses concatenation [32] of feature maps between the encoder and decoder path. In this work, the concatenation is replaced by element-wise summation (Fig. 2a). Element-wise summation directly adds the local details of the feature maps from the encoder to the global details of the feature maps from the decoder at certain stage. Thus, the residual attention modules generate attention-aware features that change adaptively as the network goes deep in layers. The novelty of this work is that the residual attention modules are added to the vanilla U-Net model [3], as in Figs. 2a and 2b, to capture multi-scale information and integrate low-level features with high-level features for precise semantic segmentation of the input MG images. By using the residual attention modules, our RU-Net can improve the performance significantly (Fig. 1). We investigated employing different number of residual attention blocks, and different number of layers. We observed that adding more than two residual blocks and four layers does not significantly improve the model's performance; but it significantly increases the training time. Therefore we considered the architecture shown in Table 2 for the proposed RU-Net model.

3.4 Cascaded Residual Classifier

The classification of benign and malignant mass lesions is one of the most challenging and also the most significant processes in examining MG images as it helps to reduce false positives (FPs) and classify the lesions at their early stage. Almost all the DL methods proposed for classifying MG images use the gray-scale images [4]. We used images of 224×224 pixels to train our DL model for classification of lesions in MG images into benign or malignant. The size 224×224 are used excessively for training DL CNNs [1, 4]. Classifying lesions into benign or malignant using gray-scale images is challenging because in some MG images, gray levels of the masses are mixed with surrounding tissues resulting in unclear lesion boundary. The augmented training binary-scale GTMs have clear mass boundaries and margins. Moreover, heterogeneously dense and dense breast tissues hide the mass lesions. These challenges decrease the accuracy of DL classifiers. In this paper, to address these challenges, we propose to use a cascaded ResNet CNN [16] trained on black and white images to classify the segmented binary maps (output of the semantic segmentation) into benign or malignant images without the need of the traditional hand-crafted features (Fig. 4).

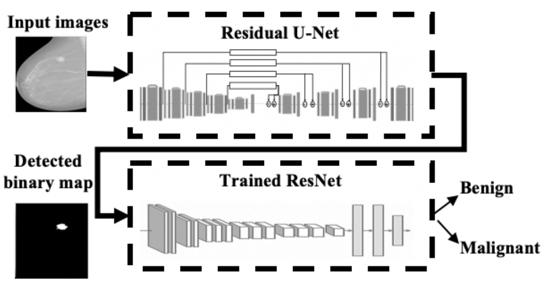


Figure 4: Block diagram of the proposed cascaded modules for semantic segmentation and classification of mass lesions.

3.5 Training Configurations

For training the segmentation models, we adopted the Dice coefficient (DC) loss [23] as the objective function to train the model. The DC loss function is minimized using Adam optimizer with a decreasing learning rate (LR) initialized to 10^{-2} and momentum of 0.9. The LR is reduced every 25 epochs by a factor of 0.1. We trained the models for 150 epochs. We trained the models using mini-batches of size 4. To manage imbalance data, we introduced class weights into the DC loss function.

We utilized the ResNet CNN model [16] in this study for classification of masses into benign or malignant. In our experiment, the ResNet model [16] is trained using a Stochastic Gradient Descent with a gamma of 0.1, and a weight-decay of 10^{-5}. We trained the models using mini-batches of size 16. We used initial LR of 10^{-3}, which is reduced every 25 epochs by a factor of 0.1. A dropout of 0.5 is used to accelerate the training process and prevent overfitting. For training the classification models, we used GTMs of the DDSM and BCDR-01 data-sets to train the model into benign or malignant images. We trained the models with 1,098 benign MG images, and 1,771 malignant MG images, respectively, as shown in Table 1. We applied the same augmentation technique discussed in the Material and Method Section to augment the binary GTMs images. We performed data augmentation to alleviate the relatively small amount of training data-set. To evaluate the performance of the ResNet model for classification, we carried out 5-fold cross-validation tests on the INbreast data-set (Table 1). The detected segmented binary images that have IOU $\geq$ 0.4 with the GTMs are used to test the cascaded classifier. The detected segmented binary images are resized to 224×224 pixels. We used a pre-trained Res-Net model trained on natural images and then, fine-tune it with binary-scale GTMs by modifying the last fully connected layers to fit our task of binary classification. We also fine-tuned the pre-trained ResNet model using gray-scale MG images to classify MG images to benign or malignant. Both ResNet models were trained under the same settings and the same augmentation technique. We developed and trained the DL algorithms using MATLAB version 2018b. Training and testing the models were done on a Tesla K40m Nvidia graphics processing unit.

3.6 Evaluation Metrics

To evaluate the performance of the DL models, the DI, also known as the F1 score, and the IOU, also known as Jaccard index, metrics are used to compare the automated predicted maps with the GTMs [12]. We mapped the class probabilities from the Softmax output to discrete class labels and used them to compute the DI and IOU metrics. As mentioned in the Related Work Section, most of the lesion detection models provide BBs for an indication of a region with an abnormality. In order to compare the performance of the proposed RU-Net model with object detection models that provides BBs, such as the Faster R-CNN and Yolo, we generated a BB around every detected lesion or segment. We used the minimum and the maximum points of x and y coordinates, which indicate the locations of masses to generate the BBs. We considered a detected segment (or a BB) as true positive (TP) if it overlaps with the ground truth segment (or BB) by more than 40%. For each class, the pixel accuracy metric is the ratio of correctly classified pixels to the total number of pixels in that class, according to the GTMs. Mean pixel accuracy is the mean accuracy of all classes in all images. We also calculated the Boundary F1 contour matching score (BF-score) for each image, which indicates how well the predicted boundary of each class aligns with the true boundary. For each class, mean BF-score shows the mean BF-score of all classes in all images, where values near 1 show perfect boundary.

4 RESULTS AND DISCUSSIONS

4.1 Segmentation Results

We compared the performance of the original U-Net model [32], vanilla U-Net [3], Faster R-CNN model [31], and Yolo model [6] in detecting masses with that of our proposed model in terms of mean DI, mean IOU, and the inference time in seconds per image. Further details about the implementation of these models can be found in the original article in [3, 6, 31, 32]. We trained the above models using our augmented data-set and tested them using the INbreast data-set [10]. The outputs of the Faster R-CNN and the Yolo models are BBs per inference, as shown in Fig. 5 (h: i) in red. These models provide multiple BBs. Overlapping BBs are merged

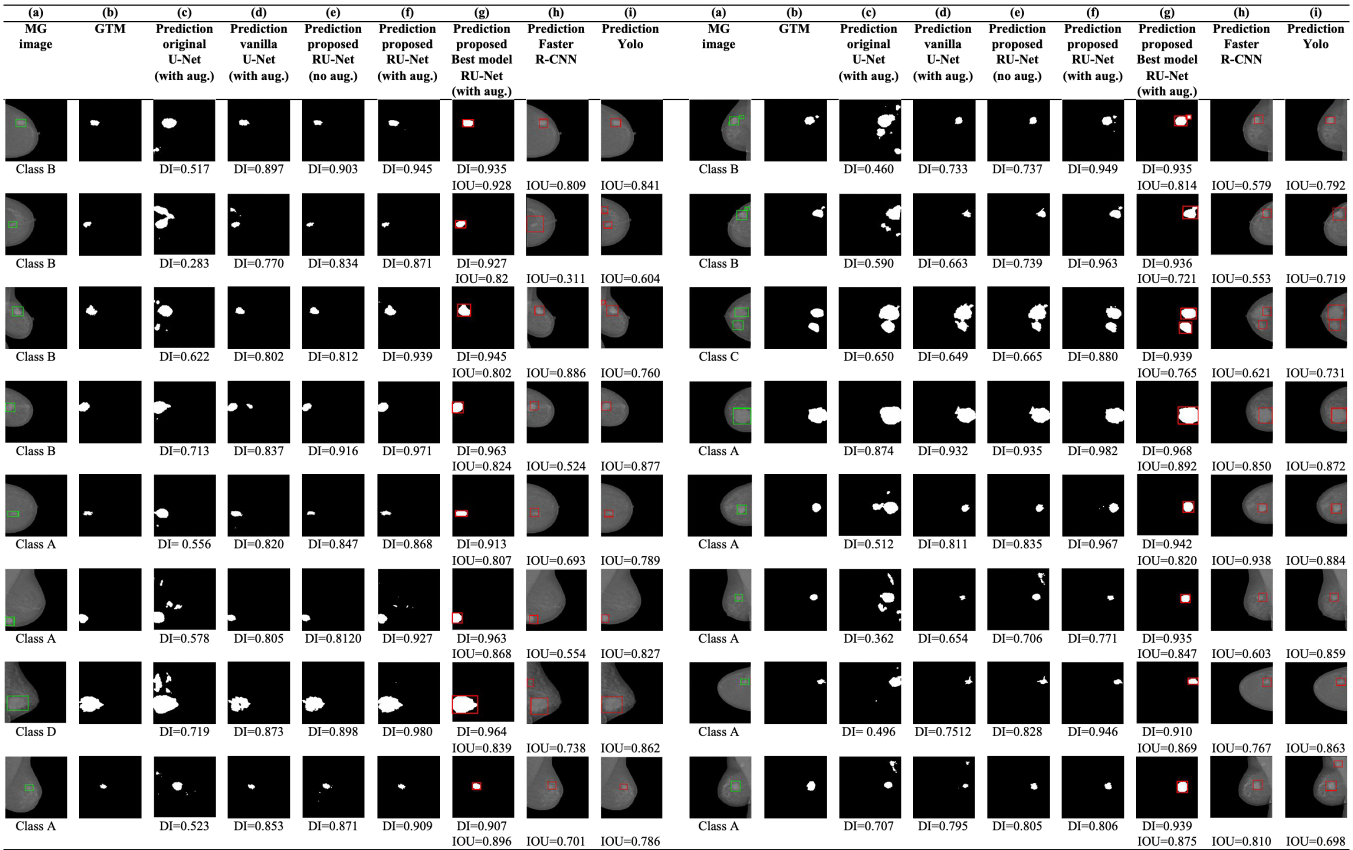


Figure 5: A comparison between the original U-Net model, vanilla U-Net model, Faster R-CNN model, Yolo model, and the proposed RU-Net model for detection of mass lesions in the INbreast database.

when the IOU between two boxes exceeds 0.5. The BB with the highest confidence score among the set of overlapping inference results is used as the representative BB with its confidence score. Figure 5a shows the original FFDM MG images from the INbreast data-set [10], where the green BBs show the location of lesions as given by experienced radiologists, (b) shows the associated pixel-level GTMs of the mass lesions, (c) shows the prediction of the original U-Net model [32], (d) shows the prediction of the vanilla U-Net model [3], (e) shows the prediction of the proposed RU-Net model trained for 150 epochs without augmentation, (f) shows the prediction of the proposed RU-Net model trained for 150 epochs on the augmented data-set, (g) shows the output of the proposed best RU-Net model trained for 200 epochs on the augmented data-set, (h) shows the prediction of the Faster R-CNN model, and finally (i) shows the prediction of the Yolo model. In Fig. 5g, we constructed red BBs that surround the predicted masses using the proposed RU-Net method in order to compare them with the Faster R-CNN and Yolo detected BBs. The calculated DI and/or IOU for each prediction is shown under each image in Fig. 5. The DI number shows the Dice similarity coefficient between the current predicted map and its corresponding GTM. DI takes a value in the range [0, 1], where 1 means that the segmentations in the two images is a perfect match. We also labeled each MG image with the corresponding breast density class (Fig. 5a). We added these labels in order to visually find the effect of breast density on detection of masses in MGs across different models. Classes A, B, C, and D are corresponding to fatty, scattered, heterogeneously dense, and dense classes, respectively.

The proposed RU-Net outperforms the segmentation results of the vanilla U-Net model [3] and the original U-Net model [32] in terms of DI (Fig. 5). Its better performance is due to the superiority of its network architecture. In the vanilla and original U-Net models, the aggregations from the encoder to decoder consist of simple and linear skip connections. As a result, the high-resolution features aggregated to the decoder are relatively shallow. Even-thought the vanilla U-Net model gives better results than the original U-Net in terms of DI, the proposed RU-Net gives more precise segmentation results. The false positive (FP) segments in some of the scattered MG images in Fig. 5 (c: d) disappeared when using the proposed RU-Net. Figure 5g shows high IOU values against the GTMs in comparison to the IOU of the Faster R-CNN and the Yolo models, Fig. 5 (h and i). However, Yolo provide better precise detection than the Faster R-CNN model due to having adaptive anchor boxes generated from the training data-set.

Detecting small lesions in MGs are very challenging, especially if these small lesions exist in heterogeneously dense and dense MG images. In order to address the problem of detecting small lesions, the proposed RU-Net model uses residual attention blocks to increase resolution for better pixel-level prediction. By using residual blocks, the network incorporates multi-scale spatial context and captures more local and global context to predict a precise pixel-wise segmentation map of an input full MG image he high-resolution features from the encoder are aggregated more for obtaining stronger semantic information (Fig. 5 (e: g)). In order to obtain more accurate segmentation results, we used large images of 640×640 for training the network to provide much contextual information. One of Yolo and Faster R-CNN known drawback is having low detection accuracy on small objects. In Figs. 5, 6 and 7, small lesions in MGs tend to have very low IOU compared to our proposed RU-Net.

The proposed RU-Net model overcomes the limitation of the state-of-the-art DL segmentation models in terms of reserving high-resolution details by using the residual attention modules, which help the model to detect masses in dense images. Figure 6 illustrates the capability of the proposed RU-Net in detecting small masses in heterogeneously dense and dense MG images. Moreover, masses that existed over the pectoral muscle in dense areas are detected by DI ≥0.65 (Fig. 6, first row). The proposed RU-Net model succeeds to detect multiple lesions in the same breast as shown in Figs. 5, 6 and 7. Moreover, the proposed RU-Net can precisely detect multiple lesions with higher DI and IOU in comparison to the other methods (Fig. 7, first row).

The Faster R-CNN and the Yolo models are able to detect lesions in the MG images, however, these models introduce more FPs (In Fig. 7, second row). The Faster R-CNN and the Yolo models also results in more false negatives (FNs), as shown in last row of Fig.7, where they provide very low IOU that reaches 0.0 with FPs, while the proposed RU-Net have IOU that ≥ 0.7. As we mentioned before, the MG images of class C (heterogeneously dense) is very challenging where the dense areas of the breast make it harder to find masses and obscure small masses. The proposed RU-Net succeed to overcome these challenges (Figs. 5, 6 and 7). This is because the segmentation process incorporates more multi-scale spatial context and captures more local and global context to predict a precise pixel-wise segmentation map of an input full MG image.

We divided the MG images in the INbreast data-set into 4 classes accordingly to its breast density classes. The numbers of images in classes A, B, C, D are 42, 36, 21, 8 images, respectively. We tested the 5 models individually on MG images in each class. Figure 8 shows the histogram of the IOU for each class, with bin width of 0.1. The RU-Net detected 100% fatty MG images with an IOU of range of 1: 0.9 (Fig. 8a). The vanilla U-Net follows the proposed network in the detection of masses with 11.9% and 78.57% of fatty MG images having an IOU in the range of 1: 0.9 and 0.9: 0.8, respectively, while the Faster R-CNN and Yolo models show lower IOU in the range of 1: 0.8 (Fig. 8b). In the case of the scattered MG images, the RU-Net and the vanilla U-Net have nearly the same IOU histogram distributions in the ranges 1: 0.8 (Fig. 8b). In the challenging cases, the heterogeneously dense and dense MG images, the proposed RU-Net is superior in detecting masses with high IOU in the range of 1: 0.8 than other models (Figs. 8c and 8d). We noticed that the Faster R-CNN have 43% of all density classes in the range of 0.5: 0.4. The RU-Net detected 53.27%, 42.9%, 2.8% of masses in MG images in all density classes with IOU in the range of 1: 0.9, 0.9: 0.8, 0.8: 0.7, respectively. While the vanilla U-Net detected 13.08%, 77.57%, 6.5% of masses in MG images in all density classes with IOU in the range of 1: 0.9, 0.9: 0.8, 0.8: 0.7, respectively (Fig. 8). We noticed that the masses that exist in the dense category in the INbreast data-set are relatively larger in terms of size than the other masses in the other categories. That is the reason that why the Faster R-CNN and Yolo models have high IOU in that category (class D) than other categories (classes A, B, C) (Fig. 8d).

In Table 3, the performance of the models under study is shown for comparison between the tight detected BBs and the ground truth BBs. Moreover, Table 3 shows a breakdown of the values of the mean accuracy and mean IOU among different breast density classes. The proposed RU-Net is superior in detecting masses than other models under study with high mean accuracy and mean IOU. The BF-score of the proposed RU-Net method is 0.981 which exceeds the other segmentation models under study. The values of DI and IOU of the vanilla U-Net method is closer to that of the RU-Net method compared to the original U-Net. The trained original U-Net has a mean DI of 0.756, a mean IOU of 0.836, respectively.

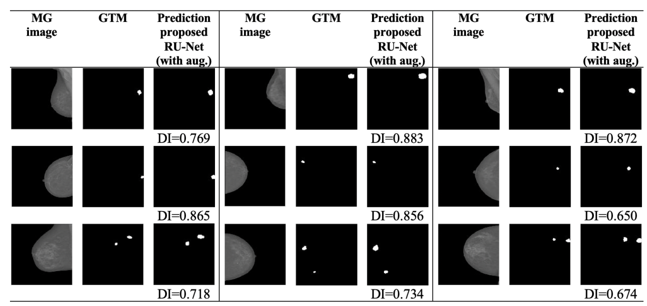


Figure 6: Shows the detection of small lesions in the pectoral muscle, small lesions in heterogeneous dense MG images, and multiple small lesions.

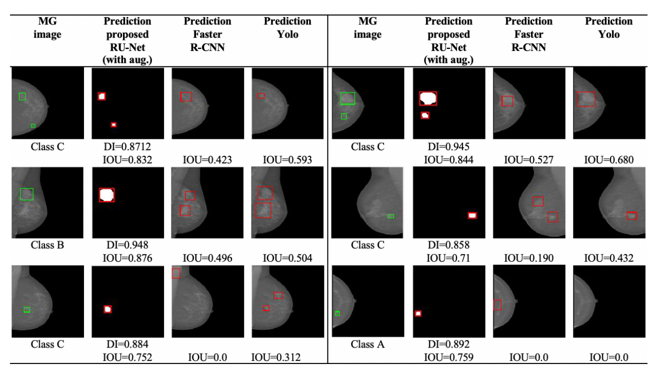
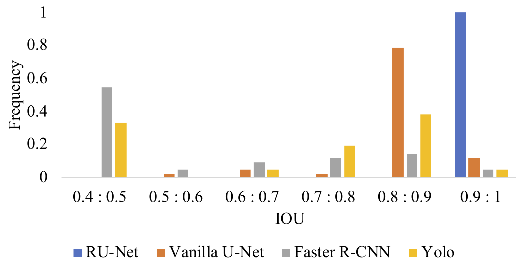
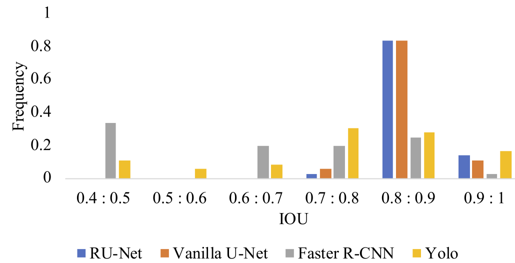


Figure 7: A comparison between the proposed RU-Net and the Faster R-CNN and Yolo models in terms of IOU.

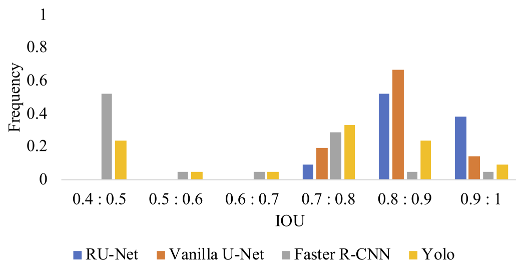
Effect of augmentation. We investigated the effect of augmentation in the performance of the proposed RU-Net method, shown in Fig. 5 (e: g). For example, the values of the DI of the augmented model, as shown in (f and g), are higher than the ones of the trained model



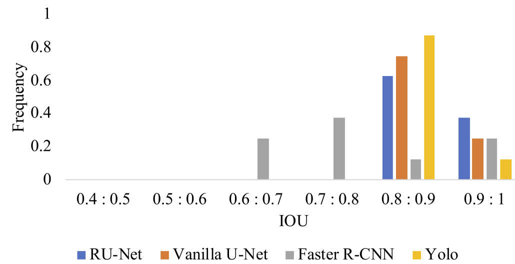
(a) Fatty MG images.



(b) Scattered MG images.



(c) Heterogeneous dense MG images.



(d) Dense MG images.

Figure 8: Histogram of the mean of IOU value for the test images in each MG breast density class.

Table 3: The performance of the vanilla U-Net, Faster R-CNN, Yolo, and the proposed RU-Net model.

Model	Class A		Class B		Class C		Class D	
	Mean acc.	Mean IOU	Mean acc.	Mean IOU	Mean acc.	Mean IOU	Mean acc.	Mean IOU
-Proposed RU-Net	0.997	0.981	0.929	0.937	0.995	0.930	0.940	0.928
-Vanilla U-Net	0.921	0.873	0.931	0.912	0.912	0.911	0.939	0.912
-Faster R-CNN	0.763	0.717	0.799	0.732	0.733	0.689	0.881	0.797
-Yolo	0.886	0.814	0.878	0.791	0.911	0.778	0.910	0.828

Table 4: The performance of the proposed RU-Net without aug., vanilla U-Net with aug., original U-Net with aug., and the proposed RU-Net model with aug.

Model	Mean acc.	Mean IOU	Mean BF-Score	Mean Dice
-Proposed RU-Net, aug.	0.987	0.948	0.981	0.983
-Proposed RU-Net, no aug.	0.944	0.891	0.919	0.905
-Vanilla U-Net [3]	0.962	0.921	0.926	0.943
-Original U-Net [32]	0.843	0.836	0.789	0.756

without augmentation, as shown in Fig. 5 (e). The mean DI and the IOU of the proposed RU-Net with augmentation are 0.983 and 0.948, respectively, compared to 0.905 and 0.891 of the RU-Net model without data-augmentation (Table 4). The BF-score improves from 0.919 to 0.981 in the case of the proposed augmented U-Net model. Figure 5 (e: g), shows that the DI per image increases when the proposed model is trained with the augmented mixed data-set.

Improvements of the proposed RU-Net model. The proposed model yields an improvement of 4.24%, 30.02% in the DI and 2.93%, 13.39% in the IOU, respectively, relative to that of the vanilla U-Net model and the original U-Net (Table 4). The original U-net architecture uses a long skip connection to concatenate the features maps. By replacing the concatenation module with an addition module, the RU-Net becomes a fully residual attention model.

Timing performance. To assess the runtime performance of these methods, we measured the inference time per image taken by each method to detect lesions in the test data-set. The mean inference time per image of the proposed RU-Net method, Vanilla U-Net, Original U-Net, Faster R-CNN, and Yolo models are 0.094, 0.087, 0.080, 0.439, 0.206 seconds, respectively. Its running time is comparable (slower in fractions of milliseconds) with the Vanilla U-Net and original U-Net, while outperforming the Faster R-CNN, and Yolo models. We have to emphasize that for radiologists, an inference time of a fraction of second or even several seconds is not as important as the accuracy of the given model.

4.2 Classification Results

We evaluated the performance of the ResNet CNN model in terms of area under the ROC curve (AUC) for the task of classifying the segmented binary maps of breast masses as benign or malignant. Our approach validates the usefulness of using binary GTMs for training the ResNet model. As we mentioned in the Evaluation Metrics subsection, the detected segmented binary images that have IOU equals or exceeds 40% comparing with its GTMs are used to test the cascaded classifier. The IOU range of 0.4: 0.6 is commonly used with DL localization models [4, 5, 7]. However, the results shown in Figs. 8a, 8b, 8c, and 8d demonstrate that our RU-Net model accurately detects masses with IOU of value higher than 0.7, which allows stable high performance in detecting lesions across different breast density categories. The classification performance is evaluated in terms of sensitivity (sen.), specificity (spe.), accuracy (acc.), F1-score, and Matthews Correlation Coefficient (MCC) per image. Tables 5 and 6 show the classification performance for the ResNet models trained on augmented binary-scale MG maps and augmented gray-scale MG images using 5-fold cross-validation. Table 6 shows the confusion matrix of the classification task.

In our experiments, we used 224×224 pixels to train the ResNet, because firstly, binary-scale lesions have less details and are well distinguishable than gray-scale images. Secondly, with transfer learning, data augmentation and the structure of the ResNet model (residual blocks), better results are achieved to classify lesions of different size and shapes (Tables 5 and 6). The ResNet model results in a TPR/FPR of 0.94/0.03 when trained on binary-scale GTMs and a TPR/FPR of 0.87/0.07 when trained on gray-scale images using the INbreast test data-set (Table 6). The accuracy of the ResNet model is 0.95 and 0.91 when trained with binary-scale GTMs images and gray-scale images, respectively. It is observed that the ResNet model that is fine-tuned with gray-scale GTMs performs better than the one fine-tuned with gray-scale MG with mean sen. (0.94), spe. (0.96), acc. (0.95), AUC (0.98), F1-score (0.93), and MCC of (0.90) (Table 5). In Table 6, 94.84% of benign cases and 96.08% of malignant cases are correctly classified, while 3.92% and 5.16% are falsely classified using the ResNet model trained on GTMs (Table 6). However, the ResNet model that is fine-tuned using gray-scale images results in 7.63% and 12.58% of miss classified benign cases and malignant cases, respectively (Table 6).

Table 5: Comparison of the classification performance of the ResNet model over 5-fold cross validation using data-set trained on augmented binary-scale GTMs images or gray-scale MG images.

Model	Sen.	Spe.	Acc.	AUC	F1 score	MCC
-Proposed ResNet, binary-scale images	0.94	0.96	0.95	0.98	0.93	0.90
-ResNet, gray-scale images	0.87	0.92	0.91	0.96	0.86	0.79

The results of classification show the robustness of the proposed ResNet CNN in minimizing the FP and FN rates (Tables 5 and 6). The improved performance of the ResNet classifier of binary images is due to the following reasons. First, the training GTMs have clear boundaries and margins. Second, the high deep level of features from proposed RU-net contributed to improving the performance of the cascaded classifier by producing precise segmentation maps that can be correctly classified as benign or malignant with high accuracy (Tables 5 and 6). In our proposed cascaded Res-Net classifier, the model classify the full detected map. The mean inference time of the classification model is 0.033 seconds.

Table 6: Confusion matrix of the classification task via the ResNet model trained on binary-scale or gray-scale MG images over 5-fold cross validation.

Model	Actual classes	Predicted classes	
		Benign	Malignant
-Proposed ResNet, binary-scale images	Benign	0.94	0.05
	Malignant	0.03	0.96
-ResNet, gray-scale images	Benign	0.87	0.12
	Malignant	0.07	0.92

5 CONCLUSIONS

We propose a novel network architecture, which consists of two cascaded convolutional neural networks (CNNs). The first network is a residual U-Net (RU-Net) for semantic segmentation of mass lesions in mammography (MG) images. The proposed RU-Net predict a pixel-wise segmentation binary map of an input full MG image in an efficient way due to the residual attention modules. The second network, a ResNet classifier, is used for the task of binary shape classification of the segmented binary-scale maps into benign or malignant. We compared the performance of the proposed RU-Net model with the performance of two DL semantic segmentation models, U-Net and Vanilla U-Net, and two DL object detector models, YOLO and Faster R-CNN. We trained all the models with the same data-sets. We observed that the data augmentation used to increase the training data-set size enhances the performance of the proposed model. We also observed that the proposed model has a lower mean inference run time per image compared to the other DL models. The proposed RU-Net model achieves a mean test pixel accuracy of 0.98, mean Dice coefficient index (DI) of 0.98 and mean intersection over union (IOU) of 0.94 that outperform those of the other models. In summary, the proposed RU-Net model can be used for precise segmentation of masses in MG images, especially for the challenging heterogeneously dense and dense breast cases. The proposed RU-Net is superior in detecting masses in dense MGs with high IOU in the range of 0.8: 1 than other models. This is because the segmentation process incorporates more multi-scale spatial context and captures more local and global context to predict a precise pixel-wise segmentation map of an input full MG image. The results show that the precise segmented masses can be used for more accurately differentiating benign from malignant lesions that is a very challenging task. The fine-tuned ResNet model with binary-scale GTMs performs better than the ResNet model fine-tuned using gray-scale MG images in terms of mean sensitivity (0.94), specificity (0.96), accuracy (0.95), AUC (0.98), F1-score (0.93), and Matthews Correlation Coefficient (0.90). To conclude, using transfer learning, introducing augmentation, and incorporating multi-scale local and global context using the residual attention modules into the original U-Net architecture result in a better performance in detecting and segmenting masses which can be used in more effective classification of masses.

ACKNOWLEDGMENTS

This work is supported by Sheida Nabavi's startup money from University of Connecticut and Dina Abdelhafiz's scholarship from the ministry of higher education and scientific research, Egypt and City of Scientific Research and Technological Applications (SRTA-City), Egypt.

REFERENCES

[1] Dina Abdelhafiz, Sheida Nabavi, Reda Ammar, and Clifford Yang. 2017. Survey on deep convolutional neural networks in mammography. In *2017 IEEE 7th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*. IEEE, 1–1. https://doi.org/10.1109/ICCABS.2017.8114310

[2] Dina Abdelhafiz, Sheida Nabavi, Reda Ammar, and Clifford Yang. 2018. The Effect of Pre-Processing on Breast Cancer Detection Using Convolutional Neural Networks. *Poster session presented at the meeting of the IEEE International Symposium on Biomedical Imaging,Washington, DC* (2018).

[3] Dina Abdelhafiz, Sheida Nabavi, Reda Ammar, Clifford Yang, and Jinbo Bi. 2018. Convolutional Neural Network for Automated Mass Segmentation in Mammography. In *2018 IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*. IEEE, 1–1.

[4] Dina Abdelhafiz, Clifford Yang, Reda Ammar, and Sheida Nabavi. 2019. Deep convolutional neural networks for mammography: advances, challenges and applications. *BMC Bioinformatics* 20, 11 (06 Jun 2019), 281. https://doi.org/10.1186/s12859-019-2823-4

[5] Mugahed A Al-antari, Mohammed A Al-masni, Mun-Taek Choi, Seung-Moo Han, and Tae-Seong Kim. 2018. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International Journal of Medical Informatics* 117 (2018), 44–54.

[6] Mohammed A Al-masni, Mugahed A Al-antari, JM Park, Geon Gi, Tae-Yeon Kim, Patricio Rivera, Edwin Valarezo, S-M Han, and T-S Kim. 2017. Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 1230–1233.

[7] Mohammed A Al-masni, Mugahed A Al-antari, Jeong-Min Park, Geon Gi, Tae-Yeon Kim, Patricio Rivera, Edwin Valarezo, Mun-Taek Choi, Seung-Moo Han, and Tae-Seong Kim. 2018. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Computer methods and programs in biomedicine* 157 (2018), 85–94.

[8] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. 2015. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561* (2015).

[9] Corinne Balleyguier, Salma Ayadi, Kim Van Nguyen, Daniel Vanel, Clarisse Dromain, and Robert Sigal. 2007. BIRADSâĎć classification in mammography. *European journal of radiology* 61, 2 (2007), 192–194.

[10] Antonio Cardoso, Jaime Cardoso, Maria Joćo Cardoso, Ines Moreira, Igor Amaral, and Inźs Campos Domingues. 2012. INbreast: Towards a Full Field Digital Mammographic Database. (2012).

[11] Yoni Choukroun, Ran Bakalo, Rami Ben-Ari, Ayelet Akselrod-Ballin, Ella Barkan, and Pavel Kisilev. 2017. Mammogram Classification and Abnormality Detection from Nonlocal Labels using Deep Multiple Instance Neural Network. (2017).

[12] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, Pablo Martinez-Gonzalez, and Jose Garcia-Rodriguez. 2018. A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing* 70 (2018), 41–65.

[13] M Jayesh George and S Perumal Sankar. 2017. Efficient preprocessing filters and mass segmentation techniques for mammogram images. In *2017 IEEE International Conference on Circuits and Systems (ICCS)*. IEEE, 408–413.

[14] Ross Girshick. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 1440–1448.

[15] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2980–2988.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

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

[18] Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Q Weinberger. 2016. Deep networks with stochastic depth. In *European Conference on Computer Vision*. Springer, 646–661.

[19] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[21] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram Van Ginneken, and Clara I Sánchez. 2017. A survey on deep learning in medical image analysis. *Medical image analysis* 42 (2017), 60–88.

[22] MA Guevara Lopez, NG Posada, Daniel C Moura, Raúl Ramos Pollán, José M Franco Valiente, César Suárez Ortega, M Solar, G Diaz-Herrero, IMAP Ramos, J Loureiro, et al. 2012. BCDR: a breast cancer digital repository. In *15th International Conference on Experimental Mechanics*.

[23] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 565–571.

[24] Arnau Oliver, Jordi Freixenet, Joan Marti, Elsa Perez, Josep Pont, Erika RE Denton, and Reyer Zwiggelaar. 2010. A review of automatic mass detection and segmentation in mammographic images. *Medical image analysis* 14, 2 (2010), 87–110.

[25] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2010), 1345–1359.

[26] Stephen M Pizer, R Eugene Johnston, James P Ericksen, Bonnie C Yankaskas, and Keith E Muller. 1990. Contrast-limited adaptive histogram equalization: speed and effectiveness. In *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*. IEEE, 337–345.

[27] Waseem Rawat and Zenghui Wang. 2017. Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation* 29, 9 (2017), 2352–2449.

[28] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

[29] Reza Reiazi, Reza Paydar, Ali Abbasian Ardakani, and Maryam Etedadialiabadi. [n. d.]. Mammography lesion detection using Faster R-CNN detector. *Computer Science & Information Technology* ([n. d.]), 111.

[30] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 6 (2017), 1137–1149.

[31] Dezső Ribli, Anna Horváth, Zsuzsa Unger, Péter Pollner, and István Csabai. 2018. Detecting and classifying lesions in mammograms with Deep Learning. *Scientific reports* 8, 1 (2018), 4165.

[32] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.

[33] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* 15, 1 (2014), 1929–1958.

[34] Jonas Teuwen, Sil van de Leemput, Albert Gubern-Mérida, Alejandro Rodriguez-Ruiz, Ritse Mann, and Babak Ehteshami Bejnordi. 2018. Soft tissue lesion detection in mammography using deep neural networks for object detection. (2018).

[35] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. 2017. Residual attention network for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3156–3164.

[36] Pengcheng Xi, Chang Shu, and Rafik Goubran. 2018. Abnormality Detection in Mammography using Deep Convolutional Neural Networks. *arXiv preprint arXiv:1803.01906* (2018).

[37] Liliang Zhang, Liang Lin, Xiaodan Liang, and Kaiming He. 2016. Is faster r-cnn doing well for pedestrian detection?. In *European conference on computer vision*. Springer, 443–457.

[38] Wentao Zhu, Xiang Xiang, Trac D Tran, Gregory D Hager, and Xiaohui Xie. 2018. Adversarial deep structured nets for mass segmentation from mammograms. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 847–850.

[39] Lian Zou, Shaode Yu, Tiebao Meng, Zhicheng Zhang, Xiaokun Liang, and Yaoqin Xie. 2019. A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis. *Computational and Mathematical Methods in Medicine* 2019 (2019).