

METHODOLOGY

Open Access



# Convolutional neural network for automated mass segmentation in mammography

Dina Abdelhafiz<sup>1,2</sup>, Jinbo Bi<sup>1</sup>, Reda Ammar<sup>1</sup>, Clifford Yang<sup>3</sup> and Sheida Nabavi<sup>1\*</sup>

From 8th IEEE International Conference on Computational Advances in Bio and medical Sciences (ICABS 2018) Las Vegas, NV, USA. 18–20 October 2018

\*Correspondence:

[sheida.nabavi@uconn.edu](mailto:sheida.nabavi@uconn.edu)

<sup>1</sup>Department of Computer Science and Engineering, University of Connecticut, 06269 Storrs, CT, USA  
Full list of author information is available at the end of the article

## Abstract

**Background:** Automatic segmentation and localization of lesions in mammogram (MG) images are challenging even with employing advanced methods such as deep learning (DL) methods. We developed a new model based on the architecture of the semantic segmentation U-Net model to precisely segment mass lesions in MG images. The proposed end-to-end convolutional neural network (CNN) based model extracts contextual information by combining low-level and high-level features. We trained the proposed model using huge publicly available databases, (CBIS-DDSM, BCDCR-01, and INbreast), and a private database from the University of Connecticut Health Center (UHC).

**Results:** We compared the performance of the proposed model with those of the state-of-the-art DL models including the fully convolutional network (FCN), SegNet, Dilated-Net, original U-Net, and Faster R-CNN models and the conventional region growing (RG) method. The proposed Vanilla U-Net model outperforms the Faster R-CNN model significantly in terms of the runtime and the Intersection over Union metric (IOU). Training with digitized film-based and fully digitized MG images, the proposed Vanilla U-Net model achieves a mean test accuracy of 92.6%. The proposed model achieves a mean Dice coefficient index (DI) of 0.951 and a mean IOU of 0.909 that show how close the output segments are to the corresponding lesions in the ground truth maps. Data augmentation has been very effective in our experiments resulting in an increase in the mean DI and the mean IOU from 0.922 to 0.951 and 0.856 to 0.909, respectively.

**Conclusions:** The proposed Vanilla U-Net based model can be used for precise segmentation of masses in MG images. This is because the segmentation process  
(Continued on next page)



(Continued from previous page)

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. These detected maps can help radiologists in differentiating benign and malignant lesions depend on the lesion shapes. We show that using transfer learning, introducing augmentation, and modifying the architecture of the original model results in better performance in terms of the mean accuracy, the mean DI, and the mean IOU in detecting mass lesion compared to the other DL and the conventional models.

**Keywords:** Mammograms (MGs), Breast cancer, Deep learning (DL), Convolutional neural networks (CNNs), Machine learning (ML), Computer-aided detection (CAD), U-Net, Vanilla U-Net, SegNet, Ground truth maps (GTMs), Detection, Semantic pixel-wise segmentation, Localization, Pre-processing, Region growing

## Background

Breast cancer is the second most common cause of cancer death among women in the United States [1]. According to the American cancer society, the female breast cancer death rate declined by 38% from its maximum in 1989 to 2014 (avoiding about 300,000 deaths) [1]. In 2012, the estimated number of deaths among females in the USA is 43,909 out of 293,353 of all cancer deaths. Moreover, in 2017, it is estimated that there will be 40,610 breast cancer deaths in the USA [1, 2]. This decline in mortality is partially due to the advances in mammography screening and conventional computer-aided diagnosis models (CAD) [3, 4]. In the last few years, deep learning (DL) models and, in particular, convolutional neural networks (CNNs) have achieved state-of-the-art performance for image classification, lesion detection for mammography [5–7], and for medical applications in general [8]. Various approaches have been proposed to further improve the accuracy of deep CNNs [6, 7].

In a recent survey [9] on conventional CAD models and DL classification models for mammograms (MGs) images, it has been shown that conventional models have limitations in classifying MG images. Recent research studies in [4, 10–12] present different conventional models to detect lesions in MG images. Most of the conventional models depend on a pre-requisite set of local hand-crafted features that cannot be generalized to work on a new data-set. Conventional CAD models consider limited feature types (e.g. texture features, shape features, and grey level intensity features), which require expert knowledge for selecting them [4, 9, 11, 12]. Poor feature extraction and selection cause challenge to build a successful classifier [4, 6, 7, 9–12]. However, the state-of-the-art CNNs, extract global features from MG images [6, 7, 13]. In CNNs, the first layers of the network capture basic coarse features such as oriented edges, corners, textures, and lines while subsequent layers construct complex structures or global features [5].

Despite the initial success of DL models for the segmentation of lesions in medical images as general, the segmentation of lesions in mammography using DL methods has not been studied thoroughly. A few studies have used a CNN-based model for lesion segmentation [14, 15] in MGs and more research need to be done in this topic [6–8]. Few studies have employed CNN-based models for lesion detection and localization [14, 16–28]. These detectors provide bounding boxes (BBs) indicating regions of interests (ROIs), not real lesion segments. The region-based CNN (R-CNN) models [29] and its faster variants, Fast R-CNN [30], and Faster R-CNN [31] have recently become more popular for

localization tasks in mammography [18–22]. Although these detectors offer compelling advantages, training R-CNN is time-consuming and memory expensive. In R-CNN [29], the whole process involves training three independent models separately without much-shared computation: 1- the CNN for feature extraction, 2- the top SVM classifier for identifying ROIs' and 3- the regression model for tightening region BBs. The R-CNN [29] uses the Selective Search method [32] to first generate initial sub-segmentations and generate candidate regions, then it uses the greedy algorithm to recursively combine similar regions into larger ones, and lastly uses the generated regions to produce the final candidate region proposals. These region proposals lower down the number of the potential BBs [18, 19].

Instead of extracting CNN feature vectors independently for each region proposal, the Fast R-CNN [30] aggregates them into one CNN forward pass over the entire image and the region proposals share this feature matrix. Then the same feature matrix is used for learning the object classifier and the BB regressor. In R-CNN and Fast R-CNN, the region proposals are created using the Selective Search method, which is a slow process that is found to be the bottleneck of the overall detection and the localization process. The Faster R-CNN [31] is a better approach that constructs a single unified model composed of region proposal network (RPN) and Fast R-CNN with shared convolutional feature layers. The RPN is a fully convolutional network (FCN) that is trained to generate region proposals, which are then used by the Fast R-CNN for detection. The time cost of generating region proposals is much smaller in the case of RPN than Selective Search, as RPN shares the most computation with the object detection network using the shared convolution layers [30, 31].

The mask R-CNN for simultaneously detecting and segmenting object instances in an image is proposed in [33]. This model extends the Faster R-CNN model by adding a branch which is a FCN for predicting an object mask in parallel with the existing branch for BB recognition. A mass detector has been refined using a cascade of R-CNN and RF classifiers and an additional stage to eliminate false positives [21].

Patch-based CNNs [16, 17, 23, 34] were also proposed to detect masses. In [16], every breast image is divided into patches, and each patch is tested with the CNN model individually. The final detection of lesions in each case is based on the overall scores of all the patches. In [25–27] the famous YOLO CNN (You Only Look Once) [35] is used for breast mass classification and localization. YOLO [35] is a single end-to-end CNN that predicts BBs and class probabilities directly from full images in one evaluation.

Recently, the FCN and its variant improved models as U-Net [36], SegNet [37], Dilated-Net [38], have yielded outstanding results for semantic segmentation of bio-medical images and natural images [6, 13, 14]. These semantic segmentation networks are based on encoding (convolutional) and decoding (de-convolutional) layers. These approaches avoid using the fully connected layers (FCLs) of CNNs to convert the image classification networks into image semantic segmentation networks.

In this study, we developed a new model based on the architecture of the semantic segmentation U-Net model [36] to precisely segment mass lesions in MG images. In the proposed architecture, we used a pre-trained encoder layers and we added batch normalization layers (BN) [39], and dropout layers [40]. U-Net [36] is an end-to-end model that takes an image, find automated features in each layer, detects, and segments breast lesion using a single model and a unified training process. We trained the proposed Vanilla U-

Net model using large public data-sets (CBIS-DDSM [41], BCDR-01 [42], and INbreast [43]). We applied data augmentation (Aug.) to the training images to present the lesions in many different sizes, positions, angles. To enhance the contrast of the MGs, we applied image pre-processing before training the proposed model. We compared the performance of the proposed segmentation model in detecting lesions with those of the state-of-the-art Faster R-CNN [18], the conventional region growing (RG) [44], FCN [45], Dilated-Net [38], original U-Net [36], and SegNet [37] models.

## Material and methods

### Databases

We conducted our experiments on four databases, CBIS-DDSM [41], INbreast [43], UCHCDM [46], and the BCDR-01 [42]. CBIS-DDSM [41] is a digitized screen-film mammography (SFM) database that is a subset of the digitized DDSM database [47] with updated lesion segmentation and BBs, and verified pathology. We used 1,696 images from the CBIS-DDSM database that have mass lesions. BCDR-D01 is an SFM repository with 64 patients and 246 MGs [42]. In total, we used 136 mass segmentation from this database to conduct our experiments. The INbreast is another public database for MGs which comprises fully field digital mammography (FFDM) images [43]. It has a total of 410 images, and we used 116 MGs that are annotated for masses. UCHCDM is a private database of FFDM images collected from the University of Connecticut health center (UCHC) [46, 48]. In total, the UCHCDM database consists of 173 patients with 1,340 FFDM images. We selected 59 cases out of the 173 that have mass lesions, with a total of 118 MGs with mass annotations. The CBIS-DDSM, INbreast, and UCHCDM data-sets include separate files that show region of interest (ROI) annotations for the abnormalities, provided by radiologists [6, 7].

We combined these databases and generated a new data-set containing MGs with different resolutions (see supplementary Fig.1, Additional file 1). This new data-set provides mass lesions of different sizes, shapes, and margins. All images containing suspicious areas have associated pixel-level ground truth maps (GTMs) indicating the true locations of suspicious regions (see supplementary Fig.2, Additional file 1). The total number of images used in this combined data-set is 2,066 and each image has its corresponding GTM. We divided the images into a training data-set of 1,714 images, validation data-set of 204 images, and test data-set of 148 images. Images reserved for testing were not used in the training and the validation data-set. Images that come from the same patient were not split across the training and test data-sets.

### Pre-processing

Pre-processing of MGs is an essential step before applying DL methods. Its main goal is to enhance the characteristics of MGs by applying a set of filters to improve the performance of the downstream analysis. First, we detect the breast boundary for removing a big portion of the black background [49, 50]. After that, we apply the adaptive median filter (AMF) [51] to remove any existing noises. Then, we employ the contrast limited adaptive histogram equalization (CLAHE) [52] to enhance the contrast of the MGs [6, 49, 50], see supplementary Pre-processing subsection, Additional file 1. The superior performance of the CLAHE filter compared to other filters are shown in [6, 53]. All full MGs are converted into png format and re-sized to 512×512.

### Data augmentation

In this study, we adopted augmentation techniques to increase the size of our training data-set to avoid overfitting the model. We adopted the augmentation techniques used in [54–56]. We generated augmented images by image rotation in a range of  $\pm 10$  degrees, left-right flips, translate images left and right by 10%, translate images up and down by 10%, and zoom in and out by 20%. The mass segmentation maps are represented by binary images that are cropped, re-sized and augmented in the same way as their corresponding MGs. All pixels in the GTMs are labeled as belonging to background or breast lesion classes. The size of the generated augmented data-set is ten times larger than the size of the original data-set.

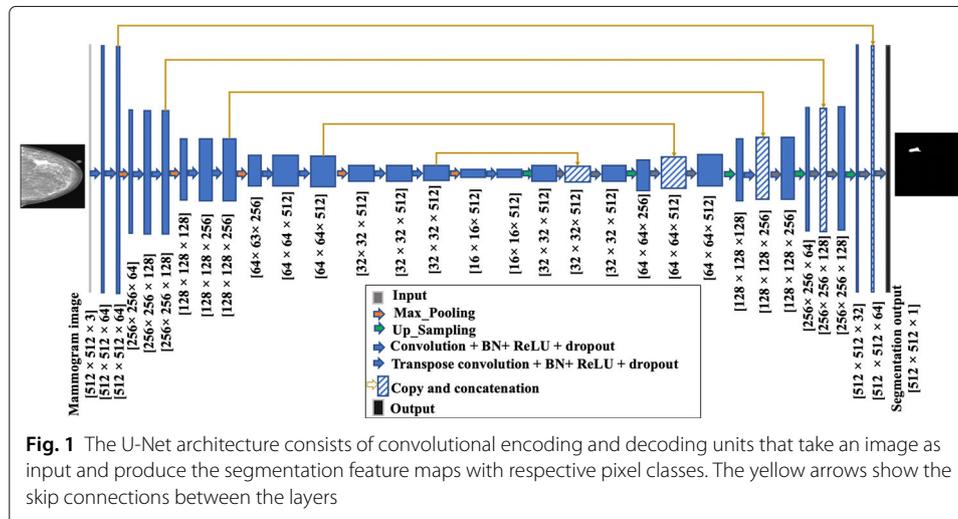
### Semantic segmentation using U-Net

The U-Net is a popular end-to-end encoder-decoder network for semantic segmentation that is originally invented for bio-medical image segmentation tasks [36]. U-Net [36] extends the FCN [45] with a U-shape architecture, which allows features from shallower layers to combine with those from deeper layers. 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 connections that connect the high-resolution features from the contracting path to the up-sampled outputs of expanding path. After collecting the required features in the encoding path, the decoding path performs nonlinear up-sampling of the feature maps before merging with the skip connections from the encoding path followed by two  $3 \times 3$  convolutions, each followed by an element-wise rectified linear unit (ReLU). The skip concatenation allows the decoder at each stage to learn back relevant features that are lost when pooled in the encoder. The final output is obtained by passing the result through a pixel-wise Softmax classifier after the last convolution layer, which independently assigns a probability to each pixel.

### Architectural modifications

We have modified the original U-Net model [36] to improve its performance for the task of segmenting lesions. We added BN layers [39], dropout layers [40], and increased the number of convolution layers. We also trained the proposed model with augmented data-set. In our implementation, we used a pre-trained VGG-16 model [57] on ImageNet as the encoder portion of the proposed Vanilla U-Net model and thus can benefit from the features created in the encoder. Studies have shown that transfer learning techniques from one domain to another are very effective to boost the performance of the current task [6, 7]. VGG-16 [57] consists of seven convolutional layers, each followed by a ReLU activation function, and five max-polling operations. The first convolutional layer of the VGG-16 model produces 64 channels and then, as the network deepens, the number of channels doubles after each max pooling operation until it reaches 512. On the following layers, the number of channels does not change. To construct the encoder part of the Vanilla U-Net, we removed the last FCLs of the VGG-16 model and replace them with two convolutional layers of 512 channels that serves as a bottleneck part of the network, connecting the encoder with the decoder.

Figure 1 shows our modified model. The encoding path consists of five convolutional layers which perform convolution with a filter bank to produce a set of feature maps. A BN layer is added between the convolution layer and the ReLU layer. Batch normalization



**Fig. 1** The U-Net architecture consists of convolutional encoding and decoding units that take an image as input and produce the segmentation feature maps with respective pixel classes. The yellow arrows show the skip connections between the layers

[39] prevents internal covariate shifts as data are filtered through the network, and it reduces the training time, prevents data overfitting, helps stack more layers, and generally increases the performance of deep CNNs. We added drop-out layers of 0.5 after each convolutional layer to help regularize the networks [40]. Following that, max-pooling with a  $2 \times 2$  window and stride 2 is performed and the resulting output is sub-sampled by a factor of 2. The max-pooling layer reduces the dimensionality of the resulting output, enabling the further collection of features. To construct the decoder, we used transposed convolutions layers that doubles the size of the feature maps while reducing the number of channels by half. The output of a transposed convolution at each level is then concatenated with an output of the corresponding part of the decoder at the same level. Also, to keep the size of the output map the same as the size of the original input MGs, a padded convolution is applied to keep the dimensions consistent across concatenation levels.

Our data-set has imbalanced data representation. In an imbalanced representation, classes are represented by significantly different numbers of pixels, which makes the learning algorithm biased towards the dominating class (i.e. breast tissues and/or background). We address this problem by introducing class weights into the Dice loss function [58]. The class weight is the ratio of the median of class frequencies computed on the entire training set divided by the class frequency [58]. This implies that the breast tissues and background class in the training set have weights smaller than the weights of the lesion class. Moreover, we applied the augmentation techniques explained in the previous sub-section, instead of applying elastic deformations as done in the original U-Net model [36].

For training, the Dice loss function was minimized using Adam optimizer [59] with a decreasing learning rate (LR) initialized to  $1e^{-2}$  and a momentum of 0.9. We used the famous early stopping technique to avoid over-fitting the model by monitoring the DI value of the validation data-set. The training of the models stops when DI is not improved every 20 epochs. Before each epoch, the training set is shuffled and every 4 mini-batch images are then picked thus ensuring that each image is used only once in an epoch. We used input MGs re-sized to  $512 \times 512$ . We developed, trained, and tested the DL models

using MATLAB version 2019b. Training and testing the models were done on a Tesla K40m Nvidia graphics processing unit.

### Evaluation metrics

To evaluate the performance of the DL models, the Dice index coefficient (DI), also known as the F1 score, and the Intersection over Union (IOU), also known as the Jaccard index, metrics are used to compare the automated predicted maps with the GTMs [60–62]. We mapped the class probabilities from the Softmax output to discrete class labels and used it to calculate the commonly used DI and IOU metrics, Eqs. 1 and 2, respectively.

$$DI = \frac{2TP}{2TP + FP + FN}, \quad (1)$$

where TP is the number of true positive pixels, FP is the number of false positives and FN is the number of false negatives.

$$IOU = \frac{TP}{TP + FP + FN}. \quad (2)$$

Dice index measures the similarity between the segmented lesions, that have irregular boundaries, and the annotated ground truth maps. IOU measures the intersection ratio between the obtained segmentation BBs and the ground truths BBs. Thus, IOU is used for localization of lesions and is best with rectangular boundaries. The output of different segmentation models might have similar IOU (lesion well localized) but with slightly different DI value that show how precise the lesions within the MG image are segmented.

The DI score gives more weight to TPs than FPs and FNs (Eq. 1). While IOU score gives more weight to TPs, FPs, and FNs (Eq. 2). Similar to DI, the IOU score ranges from 0: 1, with 0 signifying no overlap and 1 signifying perfectly overlapping segmentation [62]. Also, for each class, IOU can be calculated using the ratio of correctly classified pixels to the total number of ground truth and predicted pixels in that class (Eq. 2). The mean IOU of each class is weighted by the number of pixels in that class.

As mentioned in the “Background” section, most of the lesion detection models provide BBs for an indication of a region with an abnormality. To compare the performance of the proposed Vanilla U-Net model with detection models providing BBs such as the Faster R-CNN, a BB is generated around every detected lesion. The BBs are generated based on a minimum and maximum points of x and y coordinates, which indicate the locations of masses. We calculated the accuracy of localization by considering the detected segment and BB as TP if the center of the segment or the BB overlaps with the ground truth by more than 50%. For each class, the accuracy metric is the ratio of correctly classified pixels to the total number of pixels in that class, according to the GTMs (Eq. 3). Mean accuracy is the average accuracy of all classes in all images.

$$Accuracy = \frac{TP}{TP + FN}. \quad (3)$$

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, the mean BF-score shows the average BF-score of all classes in all images. Values near 1 means perfect boundary.

## Results

### Comparison with state-of-the-art methods

We adopted the Faster R-CNN model [18], original U-Net [36], VGG16-based FCN-8s model [45], VGG16-based SegNet model [37], Dilated-Net [38] model, and the conventional RG CAD model [4, 44] to apply to MGs for comparing their performances with that of our model in terms of mean accuracy, mean DI, mean IOU, mean BF-score, and the inference time in seconds per image (see Table 1). The architecture of these models in more detail is given in the Additional file 1. We trained the Faster R-CNN detector proposed in [18] to detect breast cancer lesions on MGs using our augmented data-set. We also implemented the RG method proposed in [44] and apply it to our MG images.

The test data-set consists of SFM and FFDM MG images. Figure 2b shows the SFM MG images from the DDSM database. Figures 3b and 4b show the original FFDM MG images from the INbreast database. Where the red BBs in Figs. 2b, 3b and 4b show the ground truth given by radiologists. The calculated DI and/or IOU for each detection is shown under each image.

Table 1 shows the evaluation metrics of all the networks included in this study in terms of mean accuracy, mean DI, mean IOU, mean, mean BF-score, and mean inference time (second)/image. In Table 1, the performance of the models is shown for the detected segments/tight BBs in comparison with the GTMs. The mean DI and the mean IOU of the proposed Vanilla U-Net are 0.951 and 0.909, respectively, which are higher compared to other models (Table 1). The BF-score of the proposed Vanilla U-Net model is 0.964 which exceeds the other segmentation models.

The architecture of the SegNet model is much closer to that of the U-Net model compared to the other segmentation models. However, the boundary of the detected regions of SegNet model is not aligned with the true boundary (Figs. 2f, 3f and 4f). The SegNet model has a BF-score of 0.822. The SegNet model performs better when detecting lesions in FFDM MGs compared to SFM MGs. In contrast, the proposed Vanilla U-Net model performs very well for both kinds of images (Figs. 2j, 3j and 4j). The proposed Vanilla U-Net shows better performance compared to SegNet. U-Net transfers the entire feature maps to the corresponding decoders and concatenates them to the up-sampled decoder feature maps, which gives precise segmentation. SegNet has much fewer trainable parameters compared to the U-Net model since the decoder layers use max-pooling indices

**Table 1** The performance of the proposed Vanilla U-Net model, original U-Net, Faster R-CNN, SegNet, Dilated-Net, FCN, and RG

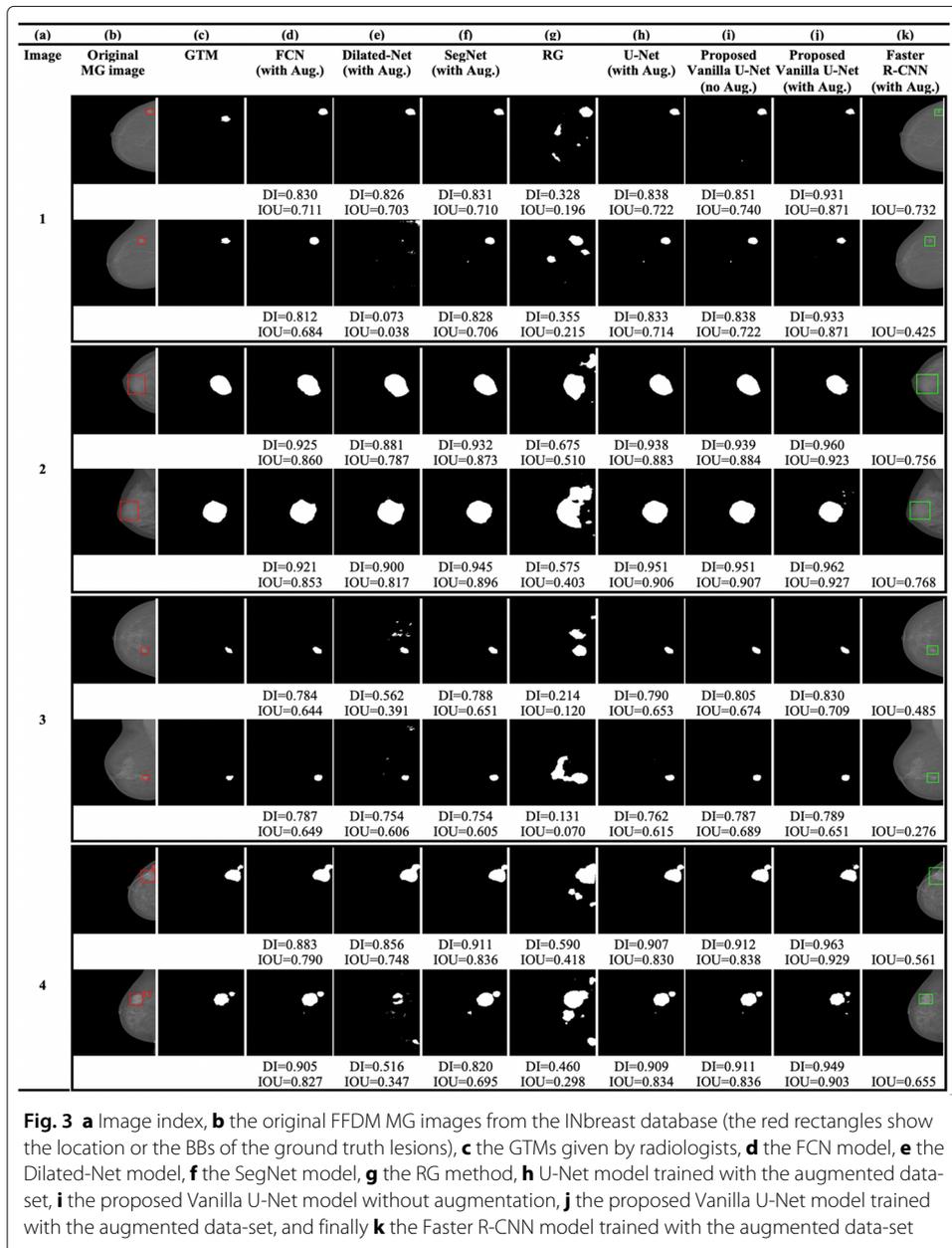
Model	Mean accuracy	Mean DI	Mean IOU	Mean BFscore	Mean inference time (second)/image
-Proposed Vanilla U-Net, with Aug.	0.926	0.951	0.909	0.964	0.115
-Proposed Vanilla U-Net, without Aug.	0.910	0.922	0.856	0.940	0.115
-Original U-Net, with Aug.	0.842	0.818	0.693	0.800	0.118
-Original U-Net, without Aug.	0.821	0.801	0.668	0.776	0.118
-Faster R-CNN, with Aug.	0.702	-	0.601	-	0.454
-SegNet, with Aug.	0.853	0.824	0.701	0.822	0.098
-Dilated-Net, with Aug.	0.832	0.799	0.665	0.701	0.094
-FCN, with Aug.	0.843	0.802	0.669	0.752	0.091
-RG	0.801	0.602	0.401	0.603	0.320

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
Image	Original MG image	GTM	FCN (with Aug.)	Dilated-Net (with Aug.)	SegNet (with Aug.)	RG	U-Net (with Aug.)	Proposed Vanilla U-Net (no Aug.)	Proposed Vanilla U-Net (with Aug.)	Faster R-CNN (with Aug.)
1										
			DI=0.406 IOU=0.255	DI=0.531 IOU=0.361	DI=0.670 IOU=0.503	DI=0.587 IOU=0.416	DI=0.648 IOU=0.479	DI=0.808 IOU=0.678	DI=0.970 IOU=0.941	IOU=0.057
2										
			DI=0.502 IOU=0.335	DI=0.612 IOU=0.441	DI=0.675 IOU=0.509	DI=0.544 IOU=0.373	DI=0.594 IOU=0.423	DI=0.770 IOU=0.627	DI=0.978 IOU=0.958	IOU=0.194
3										
			DI=0.743 IOU=0.591	DI=0.770 IOU=0.626	DI=0.759 IOU=0.611	DI=0.448 IOU=0.288	DI=0.716 IOU=0.558	DI=0.789 IOU=0.651	DI=0.940 IOU=0.887	IOU=0.381
4										
			DI=0.544 IOU=0.374	DI=0.652 IOU=0.484	DI=0.599 IOU=0.427	DI=0.355 IOU=0.216	DI=0.596 IOU=0.424	DI=0.759 IOU=0.612	DI=0.943 IOU=0.892	IOU=0.219
5										
			DI=0.882 IOU=0.789	DI=0.797 IOU=0.663	DI=0.843 IOU=0.728	DI=0.652 IOU=0.484	DI=0.885 IOU=0.794	DI=0.891 IOU=0.803	DI=0.967 IOU=0.936	IOU=0.377
6										
			DI=0.788 IOU=0.650	DI=0.575 IOU=0.404	DI=0.860 IOU=0.755	DI=0.765 IOU=0.619	DI=0.796 IOU=0.661	DI=0.828 IOU=0.706	DI=0.950 IOU=0.905	IOU=0.351
7										
			DI=0.887 IOU=0.797	DI=0.702 IOU=0.541	DI=0.888 IOU=0.798	DI=0.632 IOU=0.462	DI=0.849 IOU=0.738	DI=0.917 IOU=0.846	DI=0.968 IOU=0.938	IOU=0.467
8										
			DI=0.776 IOU=0.634	DI=0.698 IOU=0.536	DI=0.853 IOU=0.744	DI=0.527 IOU=0.358	DI=0.776 IOU=0.634	DI=0.873 IOU=0.775	DI=0.968 IOU=0.937	IOU=0.269

**Fig. 2** **a** Image index, **b** the original SFM MG images from the DDSM database (the red rectangles show the location of the BBs of the ground truth lesions), **c** the GTMs given by radiologists, **d** the FCN model, **e** the Dilated-Net model, **f** the SegNet model, **g** the RG method, **h** U-Net model trained with the augmented data-set, **i** the proposed Vanilla U-Net model without augmentation, **j** the proposed Vanilla U-Net model trained with the augmented data-set, and finally **k** the Faster R-CNN model trained with the augmented data-set

from corresponding encoder layers to perform sparse upsampling. This reduces the inference time at the decoder expanding path since the generated encoder feature maps are not involved in the upsampling. Thus, the SegNet model reveals a trade-off between the memory versus accuracy involved in achieving good segmentation performance (Table 1). The mean DI and the IOU of the trained SegNet model on augmented data-set are 0.824 and 0.701, respectively, compared to 0.952 and 0.909 of the U-Net model.

The trained Dilated-Net has a mean DI of 0.799, a mean IOU of 0.665, respectively. Moreover, its BF-score is 0.701 that is lower than that of the proposed Vanilla U-Net model and the SegNet model BF-score. Also, the performance of the Dilated-Net model is worse in the case of SFM images (Fig. 2e). Even-though some images in Figs. 2e, 3e and 4e show slightly better DI than that of SegNet, the performance of the model on all the



test data-set is lower than that of the SegNet model. In contrast to U-Net and SegNet, down-sampling layers are not required in the Dilated-Net to obtain large receptive fields and hence, high-resolution maps can be directly predicted by the model. Down-sampling layers are widely used for maintaining invariance and controlling overfitting of the model, however it reduces the spatial resolution. To retrieve the lost spatial information, the Up-sampling layers in U-Net and SegNet are used, but with additional memory and time constraints.

We also adapted the FCN-8s VGG16 based network [45] to compare its performance with that of the proposed Vanilla U-Net model. FCN-8s up-samples the final feature map by a factor of 8 after fusing feature maps from the third and fourth max-pooling layers, thus having better segmentation than its variants. The FCN in our study has a mean

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
Image	Original MG image	GTM	FCN (with Aug.)	Dilated-Net (with Aug.)	SegNet (with Aug.)	RG	U-Net (with Aug.)	Proposed Vanilla U-Net (no Aug.)	Proposed Vanilla U-Net (with Aug.)	Faster R-CNN (with Aug.)
1										
			DI=0.696 IOU=0.534	DI=0.702 IOU=0.541	DI=0.869 IOU=0.768	DI=0.090 IOU=0.047	DI=0.816 IOU=0.689	DI=0.875 IOU=0.778	DI=0.876 IOU=0.779	IOU=0.278
2										
			DI=0.809 IOU=0.679	DI=0.761 IOU=0.614	DI=0.818 IOU=0.692	DI=0.112 IOU=0.059	DI=0.819 IOU=0.693	DI=0.821 IOU=0.696	DI=0.833 IOU=0.714	IOU=0.239
3										
			DI=0.885 IOU=0.793	DI=0.580 IOU=0.409	DI=0.889 IOU=0.798	DI=0.343 IOU=0.207	DI=0.895 IOU=0.811	DI=0.894 IOU=0.808	DI=0.899 IOU=0.816	IOU=0.480
4										
			DI=0.834 IOU=0.715	DI=0.613 IOU=0.442	DI=0.836 IOU=0.718	DI=0.142 IOU=0.077	DI=0.810 IOU=0.680	DI=0.845 IOU=0.732	DI=0.847 IOU=0.734	IOU=0.117
5										
			DI=0.887 IOU=0.797	DI=0.824 IOU=0.700	DI=0.935 IOU=0.878	DI=0.382 IOU=0.236	DI=0.895 IOU=0.811	DI=0.939 IOU=0.886	DI=0.940 IOU=0.887	IOU=0.701
6										
			DI=0.890 IOU=0.802	DI=0.904 IOU=0.825	DI=0.946 IOU=0.897	DI=0.385 IOU=0.239	DI=0.904 IOU=0.825	DI=0.947 IOU=0.900	DI=0.949 IOU=0.903	IOU=0.758
7										
			DI=0.933 IOU=0.875	DI=0.928 IOU=0.865	DI=0.929 IOU=0.868	DI=0.456 IOU=0.295	DI=0.928 IOU=0.865	DI=0.941 IOU=0.889	DI=0.933 IOU=0.875	IOU=0.451
8										
			DI=0.907 IOU=0.831	DI=0.921 IOU=0.853	DI=0.930 IOU=0.869	DI=0.504 IOU=0.337	DI=0.929 IOU=0.867	DI=0.931 IOU=0.870	DI=0.943 IOU=0.892	IOU=0.696

**Fig. 4** **a** Image index, **b** the original FFDM MG images from the INbreast database (the red rectangles show the location or the BBs of the ground truth lesions), **c** the GTMs given by radiologists, **d** the FCN model, **e** the Dilated-Net model, **f** the SegNet model, **g** the RG method, **h** U-Net model trained with the augmented data-set, **i** the proposed Vanilla U-Net model without augmentation, **j** the proposed Vanilla U-Net model trained with the augmented data-set, and finally **k** the Faster R-CNN model trained with the augmented data-set

DI of 0.802 and a mean IOU of 0.669, respectively. Moreover, the BF-score of the best trained FCN is 0.752 which is lower than that of the proposed Vanilla U-Net by 0.212. The mean DI scores of the Dilated-Net and the FCN model are close for some of the images, however, the FCN give the lowest scores among all the segmentation DL models.

As we mentioned in the “Method” section, we generated tight BBs surrounding detected segments to compare the performance of the proposed model with that the BB-based models such as Faster R-CNN. The proposed Vanilla U-Net model shows better performance in detecting true segments compared to the Faster R-CNN model as shown in Figs. 2i: k, 3i: k and 4i: k. In Fig. 2, the Faster R-CNN model introduces FPs in the SFM images as in rows (1, 2, and 4 (k)). In Fig. 4k, the Faster R-CNN model introduces some FPs as in row (2k and 4k), as an example. The proposed Vanilla U-Net model shows better

performance with both FFDM and SFM images than the Faster R-CNN model. To have a better understanding of the performance of the proposed Vanilla U-Net and other models, we included the DI and/or the IOU under every image. We also show the detection of the proposed Vanilla U-Net for every CC and MLO view of the same patient. The IOU of the proposed model exceeds the IOU of the Faster R-CNN by 0.308, as shown in Table 1. We considered the detected BBs as TP if the center of the detected BB overlaps with the ground truth BB with greater than 50%.

The accurate automated seed selection process is very important for lesion segmentation. As RG segmentation's results are sensitive to the initial seed pixels, the final segmentation results would be incorrect if the seeds are not properly selected by the automated process. The RG method works better when it is used with patches of images that contain the ROI because the initial seed pixels are close to the center of the ROI. Figures 2g, 3g and 4g show the detection using the RG method. Figures 2g, 3g and 4g show that the DL models outperform the conventional CAD models in terms of DI in the segmentation of tumors in whole images. The mean DI and mean IOU of the RG method are 0.602 and 0.401, respectively.

We also explored the current state-of-the-art DL models for segmentation or localization of lesions in MG images through a literature survey [6]. The reported performance metrics of several models are shown in Supplementary Table 1, Additional file 1. The researchers used various metrics to report their work, which makes a direct comparison between these different approaches difficult. Moreover, the number of training dataset and the size of training images vary from a study to another one. However, it gives us some insights into the strategies used in these studies. Researchers who applied the transfer learning (TL) strategy to train their DL models reported that the TL approach helped them to report better detection accuracy, see supplementary Table 1, Additional file 1. Moreover, the size of the training date-set and the resolution of the MGs play an important role in increasing the model' accuracy [63].

### Effect of augmentation

In our experiments, we observed that the mean DI of the proposed Vanilla U-Net model increased slightly from that of the original U-Net model when we added BN layers or used dropout layers or increased the number of convolution layers, one at a time. Moreover, we observed that the proposed modifications, together, have increased the mean DI of the proposed Vanilla U-Net model in comparison with that of the original U-Net model from 0.801 to 0.951. But mostly the augmentation of the data-set had a great impact on the performance of the proposed Vanilla model in terms of mean DI. And because of that, we investigated the effect of augmentation in the performance of the proposed Vanilla U-Net model.

The augmented training data-set results in 17,140 images. Figures 2i: j, 3i: j, and 4i: j, illustrate the effect of augmentation on the proposed Vanilla U-Net model. For example, the values of the DI of the augmented model, as shown in (j), are higher than the ones of the trained model without augmentation, as shown in (i). Table 2 shows the improvement in terms of DI for both training and validation data-sets when using augmented data-set compared to when using the original one. The DI improves from 0.910 (training), and 0.842 (validation) to 0.972 (training), and 0.942 (validation). The augmented data-set also

**Table 2** The performance of the proposed Vanilla U-Net model

Model	Training		Validation		Test
	Dice	Loss	Dice	Loss	Dice
-Proposed Vanilla U-Net, with Aug.	0.972	0.016	0.942	0.024	0.951
-Proposed Vanilla U-Net, no Aug.	0.910	0.064	0.842	0.144	0.922

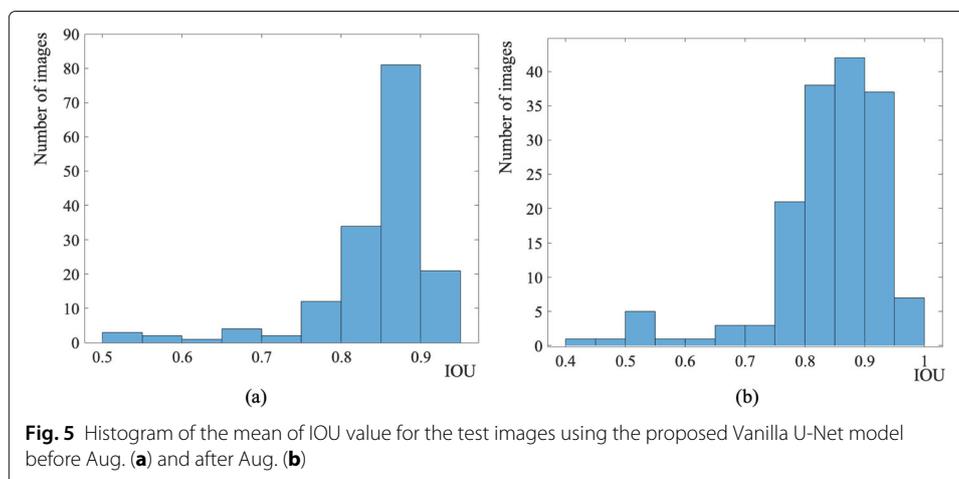
affect the localization precision significantly (Table 1). The BF-score improves from 0.940 to 0.964 in the case of the proposed augmented U-Net model.

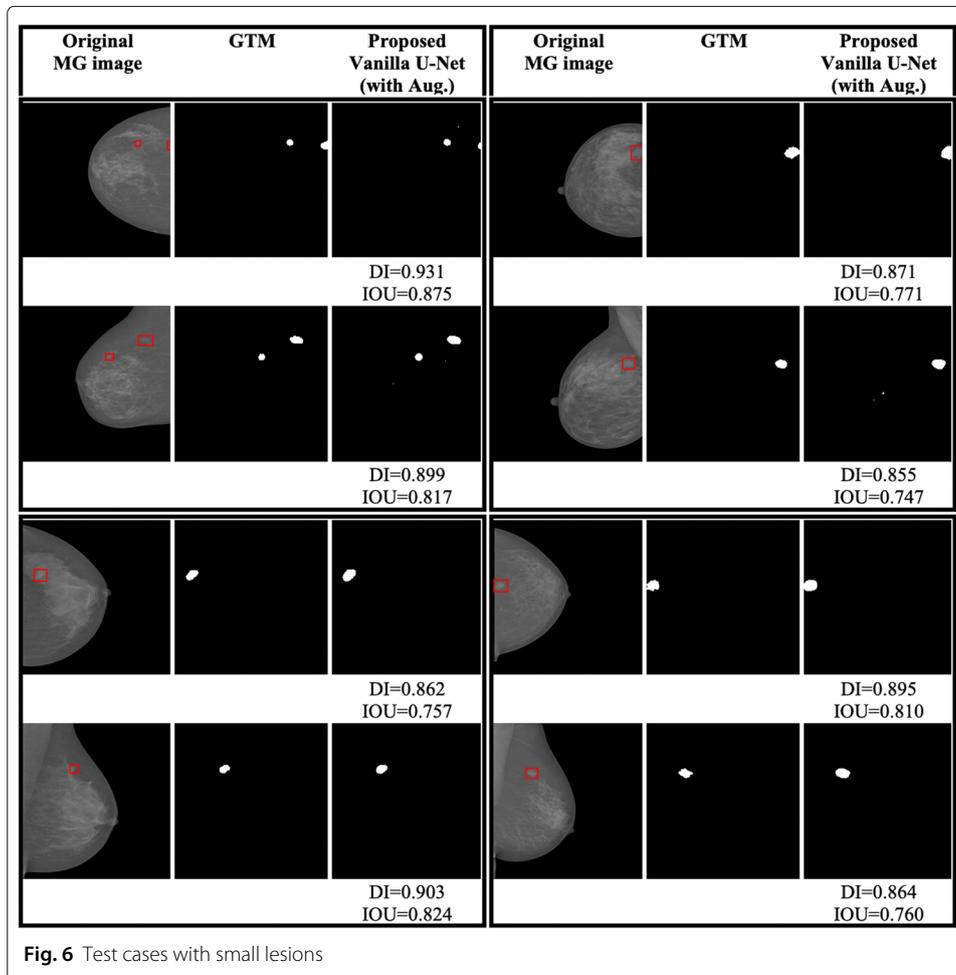
Figure 5 shows that the histogram of the mean of IOU value for the test images increases using the proposed Vanilla U-Net model after data augmentation. The mean of IOUs of the proposed Vanilla U-Net improves from 0.856 to 0.909 when training with the augmented data-set (Fig. 5 and Table 1). In general, the performance of the DL techniques improves as the size of the training data-set increases [6, 7]. Figures 2i: j, 3i: j, and 4i: j, show that the DI per image increases when the proposed model is trained with the augmented mixed data-set. The FP pixels decreased in the case of the augmented model, as shown in rows (1, 2, and 4) in Fig. 2i: j.

**Effect of image size and data-set size**

One of the factors that make a localization model or a semantic segmentation model superior to other models, is its ability to help the radiologists to detect small lesions that can be missed with the naked eye. A recent study in [63] on MGs shows that the resolution of the training images affects the performance of the CNN model. Recent studies, as shown in Table 1, Additional file 1, use MGs of small sizes as 40×40 and 227×227. The standard image sizes of 224×224, and 227×227 are used excessively for training CNNs to detect objects in natural images [61]. However, the requirement to find small mass lesions in aggressively down-sampled high-resolution images is unlikely to be successful for MGs [6, 63].

In our initial work in [14], we trained the proposed model with images of size 256×256 and found that the proposed model failed to find small lesions in images of high density. As a result, we changed our training strategy to include MGs of size 512×512 instead of 256×256. Figure 6 shows some FFDM test images that have small lesions that are detected with DI greater than 50%.





Because of the architecture of the proposed Vanilla model, images with a side divisible by 32 (e.g.  $1024 \times 1024$ ) can be used as an input to the current network implementation. In the future, we will conduct our experiments on high-resolution images to get a competitive performance to recent state-of-the-art models as the size of MG images in the clinical settings are generally larger than  $1024 \times 1024$  [7].

#### Improvements of the proposed model over the original U-Net model

The proposed model yields an improvement of 16.32% in the mean DI and 31.16% in the mean IOU, respectively, relative to that of the original U-Net model (Table 1). The original U-Net model is trained from scratch. Moreover, increasing the data-set size by using the proposed augmentation technique improves the segmentation's quality (BF-score yields an increase of 20.5% relative to that of the original U-Net model). The original U-Net did not use the BN technique. Original U-Net did not use the BN technique. Batch normalization helps the proposed model avoiding vanishing gradient problem, stacking more layers, accelerating training, and using less number of epochs. In the proposed model, we went deeper into the number of layers from four to five convolution layers. By increasing the number of convolution layers, the segmentation process incorporates more multi-scale spatial context and captures more local and global context.

### Timing performance

To assess the runtime performance of these models, we measured the mean inference time per image taken by each model to detect lesions in the test data-set, as shown in Table 1. The proposed Vanilla U-Net model is faster by 0.34 seconds than the Faster R-CNN model [18]. The inference time of the SegNet, Dilated-Net, and FCN is less than the proposed Vanilla U-Net by a fraction of second. Even though the inference time of the RG method is of about 0.3 seconds, it introduces a lot of FPs when tested on whole images as shown in Figs. 2g, 3g, and 4g, and the statistics of Table 1. The proposed Vanilla U-Net model is faster than the Faster R-CNN proposed in [19] and [20], the R-CNN proposed in [21] and [22], and the YOLO model proposed in [27], while proving a high DI, see supplementary Table 1, Additional file 1. We have to emphasize that for radiologists, the accuracy of the proposed CAD or DL model in detecting lesions is the most important feature in the mammography analysis, and the inference time is secondary. An inference time of a fraction of second or even several seconds is not as important as the accuracy of the given model.

### Discussion

We tested our proposed model on SFM and FFDM data-sets for the semantic segmentation of mass lesions in MGs. For our future work, we will consider training the proposed Vanilla U-Net model to detect both the micro-calcification and the mass lesions. We will focus on reducing FP pixels by collecting more data-sets and use higher resolution mammogram images. Finally, we want to use the proposed Vanilla U-Net model to distinguish between benign and malignant breast tumors in mammography images by studying the features of the tumors' segmented regions only.

### Conclusions

We developed a new deep learning (DL) model called Vanilla U-Net, based on the architecture of the semantic segmentation U-Net model to precisely segment mass lesions in mammogram (MG) images. The proposed end-to-end model extracts low-level and high-level features from MG images. The proposed Vanilla U-Net model efficiently predicts a pixel-wise segmentation map of an input full MG due to its modified architecture. We tested our proposed Vanilla U-Net model using film-based and fully-digital MGs. We compared the performance of our proposed model with state-of-the-art DL models namely Faster R-CNN, SegNet, FCN, and Dilated-CNN. We also compared the performance of the proposed model with the conventional region growing method. The proposed Vanilla U-Net model is superior to the segmentation models under study. The proposed Vanilla U-Net model gives a mean intersection over union (IOU) of 0.909 and a mean accuracy of 0.926 while the Faster R-CNN model gives IOU of 0.601 and a mean accuracy of 0.702, respectively. Similar to the Faster R-CNN model, the Vanilla U-Net model is trained on the full MGs. However, the proposed Vanilla U-Net model is faster and runs 0.337 seconds less than the Faster R-CNN model. We show that the proposed model show improvement in the Dice index (DI) and the IOU by 16.3% and 31.16%, respectively, relative to the original model. The proposed models can be further trained to detect micro-calcification in the future. The presented work is a step towards a precise segmentation of mass lesions in mammography. As medical data-sets are increasing and

becoming publicly available, future architectures may be trained end-to-end, removing the need for pre-training on non-medical data-sets.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12859-020-3521-y>.

**Additional file 1:** Supplementary materials (semantic segmentation using FCN, semantic segmentation using SegNet, semantic segmentation using Dilated-Net, localization using Faster R-CNN, comparison between state-of-the-art DL models, Supplementary Table 1, Supplementary Figures 1–4).

### Abbreviations

DL: Deep learning; MG: Mammogram; CNNs: Convolutional neural networks; CAD: Computer-aided detection; ML: Machine learning; TL: Transfer learning; RG: Region growing; SVM: Support vector machine; DDSM: Digital database for screening mammography; ROIs: Region of interests; GTMs: Ground truth maps; BCDR: Breast cancer digital repository; BN: Batch normalization; ReLU: Rectified liner unit; SFM: Screen-film mammography; FFDM: Digital mammography; UCHCDM: University of Connecticut health center digital mammogram; E2E: End-to-end; CLAHE: Contrast limited adaptive histogram equalization; AMF: Adaptive median filter; R-CNN: Region-based convolutional neural network; YOLO: You only look once; RPN: Region proposal network; AUC: Area under the receiver operating curve; DI: Dice index; ACC: Accuracy; IOU: Intersection over union; TP: True positive; FN: False negative; TN: True negative; FP: False positive; FPR: False positive rate; TPR: True positive rate; Aug: Augmentation; FCL: Fully connected layer; FCN: Fully convolutional network

### Acknowledgements

Not applicable.

### About this supplement

This article has been published as part of *BMC Bioinformatics Volume 21 Supplement 1, 2020: Selected articles from the 8th IEEE International Conference on Computational Advances in Bio and medical Sciences (ICCABS 2018): bioinformatics*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-21-supplement-1>.

### Authors' contributions

DH and SN designed the study. DH implemented the method, performed the experiments, and interpreted the results. DH and SN wrote the manuscript. SN supervised the study. CY provided the UCHCDM data-set. RA, CY, and JB provided feedback and helped shape the research. All authors read and approved the final version of the manuscript.

### Funding

The study is supported by the Sheida Nabavi's startup fund and Dina Abdelhafiz's scholarship from the ministry of higher education and scientific research, Egypt and the City of Scientific Research and Technological Applications (SRTA-City), Egypt. Publication costs are funded by the Sheida Nabavi' startup fund. The funding body had no influence in the design of the study, collection, analysis and interpretation of data and in writing the manuscript.

### Availability of data and materials

The DDSM data-set is available online at <http://www.eng.usf.edu/cvprg/Mammography/Database.html>. The INbreast data-set can be requested online at [http://medicalresearch.inescporto.pt/breastresearch/index.php/Get\\_INbreast\\_Database](http://medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database). The breast cancer digital repository (BCDR) data-set can be requested online at <https://www.bcdr.eu>.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Computer Science and Engineering, University of Connecticut, 06269 Storrs, CT, USA. <sup>2</sup>The Informatics Research Institute (IRI), City of Scientific Research and Technological Applications (SRTA-City), Alexandria, Egypt.

<sup>3</sup>Departments of Diagnostic Imaging, University of Connecticut Health Center, 06030 Farmington, CT, USA.

Received: 15 April 2020 Accepted: 29 April 2020

Published: 9 December 2020

### References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin.* 2017;67(1):7–30. <https://doi.org/10.3322/caac.21387>.

2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *Int J Cancer*. 2015;136(5): <https://doi.org/10.1002/ijc.29210>.
3. Souza FH, Wendland EM, Rosa MI, Polanczyk CA. Is full-field digital mammography more accurate than screen-film mammography in overall population screening? a systematic review and meta-analysis. *Breast*. 2013;22(3):217–24.
4. Tang J, Rangayyan RM, Xu J, El Naqa I, Yang Y. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Trans Inf Technol Biomed*. 2009;13(2):236–51.
5. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*; 2012. p. 1097–105. <https://doi.org/10.1145/3065386>.
6. Abdelhafiz D, Nabavi S, Ammar R, Yang C. Survey on deep convolutional neural networks in mammography. In: *Computational Advances in Bio and Medical Sciences (ICCBS)*, 2017 IEEE 7th International Conference On. IEEE; 2017. p. 1–1. <https://doi.org/10.1109/iccabs.2017.8114310>.
7. Abdelhafiz D, Yang C, Ammar R, Nabavi S. Deep convolutional neural networks for mammography: advances, challenges and applications. *BMC Bioinformatics*. 2019;20(11):281.
8. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JA, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *arXiv preprint*. 2017. arXiv:1702.05747.
9. Nahid A-A, Kong Y. Involvement of machine learning for breast cancer image classification: a survey. *Comput Math Methods Med*. 2017;2017: <https://doi.org/10.1155/2017/3781951>.
10. Liu X, Tang J. Mass classification in mammograms using selected geometry and texture features, and a new svm-based feature selection method. *IEEE Syst J*. 2013;8(3):910–20.
11. Song E, Jiang L, Jin R, Zhang L, Yuan Y, Li Q. Breast mass segmentation in mammography using plane fitting and dynamic programming. *Acad Radiol*. 2009;16(7):826–35.
12. Liu J, Chen J, Liu X, Chun L, Tang J, Deng Y. Mass segmentation using a combined method for cancer detection. *BMC Syst Biol*. 2011;5(53):6.
13. Hu Z, Tang J, Wang Z, Zhang K, Zhang L, Sun Q. Deep learning for image-based cancer detection and diagnosis—a survey. *Pattern Recogn*. 2018. <https://doi.org/10.1016/j.patcog.2018.05.014>.
14. Abdelhafiz D, Nabavi S, Ammar R, Yang C, Bi J. Convolutional neural network for automated mass segmentation in mammography. In: *2018 IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCBS)*. IEEE; 2018. p. 1. <https://doi.org/10.1109/iccabs.2018.8542071>.
15. Teuwen J, van de Leemput S, Gubern-Mérida A, Rodríguez-Ruiz A, Mann R, Bejnordi BE. Soft tissue lesion detection in mammography using deep neural networks for object detection. In: *Proceedings of the 1st conference on medical imaging with deep learning*. 2018 presented at: MIDL'18. Amsterdam; 2018. p. 1–9.
16. Sun W, Tseng T-LB, Zheng B, Qian W. A preliminary study on breast cancer risk analysis using deep neural network. In: *International Workshop on Digital Mammography*. Springer; 2016. p. 385–91. [https://doi.org/10.1007/978-3-319-41546-8\\_48](https://doi.org/10.1007/978-3-319-41546-8_48).
17. Choukroun Y, Bakalo R, Ben-Ari R, Akselrod-Ballin A, Barkan E, Kisilev P. Mammogram classification and abnormality detection from nonlocal labels using deep multiple instance neural network. In: *Eurographics workshop on visual computing for biology and medicine*. 2017.
18. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with deep learning. *Sci Rep*. 2018;8(1):4165.
19. Akselrod-Ballin A, Karlinsky L, Alpert S, Hasoul S, Ben-Ari R, Barkan E. A region based convolutional network for tumor detection and classification in breast mammography. In: *Deep Learning and Data Labeling for Medical Applications*. Springer; 2016. p. 197–205. [https://doi.org/10.1007/978-3-319-46976-8\\_21](https://doi.org/10.1007/978-3-319-46976-8_21).
20. Akselrod-Ballin A, Karlinsky L, Hazan A, Bakalo R, Horesh AB, Shoshan Y, Barkan E. Deep learning for automatic detection of abnormal findings in breast mammography. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer; 2017. p. 321–9. [https://doi.org/10.1007/978-3-319-67558-9\\_37](https://doi.org/10.1007/978-3-319-67558-9_37).
21. Dhungel N, Carneiro G, Bradley AP. The automated learning of deep features for breast mass classification from mammograms. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2016. p. 106–14. [https://doi.org/10.1007/978-3-319-46723-8\\_13](https://doi.org/10.1007/978-3-319-46723-8_13).
22. Dhungel N, Carneiro G, Bradley AP. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Med Image Anal*. 2017;37:114–28.
23. Xi P, Shu C, Goubran R. Abnormality detection in mammography using deep convolutional neural networks. *arXiv preprint*. 2018. arXiv:1803.01906.
24. Zhu W, Xiang X, Tran TD, Hager GD, Xie X. Adversarial deep structured nets for mass segmentation from mammograms. In: *Biomedical Imaging (ISBI 2018)*, 2018 IEEE 15th International Symposium On. IEEE; 2018. p. 847–50. <https://doi.org/10.1109/isbi.2018.8363704>.
25. Al-antari MA, Al-masni MA, Choi M-T, Han S-M, Kim T-S. A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification. *Int J Med Inf*. 2018;117: 44–54.
26. Al-masni MA, Al-antari MA, Park J, Gi G, Kim T-Y, Rivera P, Valarezo E, Han S-M, Kim T-S. 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; 2017. p. 1230–3. <https://doi.org/10.1109/embc.2017.8037053>.
27. Al-masni MA, Al-antari MA, Park J-M, Gi G, Kim T-Y, Rivera P, Valarezo E, Choi M-T, Han S-M, Kim T-S. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolo-based cad system. *Comput Methods Progr Biomed*. 2018;157:85–94.
28. Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, den Heeten A, Karssemeijer N. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal*. 2017;35:303–12.

29. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2014. p. 580–7. <https://doi.org/10.1109/cvpr.2014.81>.
30. Girshick R. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision; 2015. p. 1440–8.
31. Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell.* 2017;6:1137–49. <https://doi.org/10.1109/tpami.2016.2577031>.
32. Uijlings JR, Van De Sande KE, Gevers T, Smeulders AW. Selective search for object recognition. *Int J Comput Vis.* 2013;104(2):154–71.
33. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: Computer Vision (ICCV), 2017 IEEE International Conference On. IEEE; 2017. p. 2980–8.
34. Dhungel N, Carneiro G, Bradley AP. Deep learning and structured prediction for the segmentation of mass in mammograms. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2015. p. 605–12. <https://doi.org/10.1109/icip.2015.7351343>.
35. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 779–88. <https://doi.org/10.1109/cvpr.2016.91>.
36. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention. Springer; 2015. p. 234–41. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
37. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint. 2015. arXiv:1511.00561.
38. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv preprint. 2015. arXiv:1511.07122.
39. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint. 2015. arXiv:1502.03167.
40. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15(1):1929–58.
41. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *J Digit Imaging.* 2013;26(6):1045–57.
42. Lopez MG, Posada N, Moura DC, Pollán RR, Valiente JMF, Ortega CS, Solar M, Diaz-Herrero G, Ramos I, Loureiro J, et al. Bcdr: a breast cancer digital repository. In: 15th International Conference on Experimental Mechanics; 2012.
43. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. Inbreast: toward a full-field digital mammographic database. *Acad Radiol.* 2012;19(2):236–48.
44. Melouah A. Comparison of automatic seed generation methods for breast tumor detection using region growing technique. In: IFIP International Conference on Computer Science and Its Applications. Springer; 2015. p. 119–28. [https://doi.org/10.1007/978-3-319-19578-0\\_10](https://doi.org/10.1007/978-3-319-19578-0_10).
45. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015. p. 3431–40. <https://doi.org/10.1109/cvpr.2015.7298965>.
46. Zheng Y, Yang C, Merkulov A, Bandari M. Early breast cancer detection with digital mammograms using haar-like features and adaboost algorithm. In: Sensing and Analysis Technologies for Biomedical and Cognitive Applications 2016, vol. 9871. International Society for Optics and Photonics; 2016. p. 98710. <https://doi.org/10.1117/12.2227342>.
47. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer P. The digital database for screening mammography. *Digit Mammography.* 2000;431–4.
48. Zheng Y, Yang C, Merkulov A. Breast cancer screening using convolutional neural network and follow-up digital mammography. In: Computational Imaging III, vol. 10669. International Society for Optics and Photonics; 2018. p. 1066905. <https://doi.org/10.1117/12.2304564>.
49. Ramani R, Vanitha NS, Valarmathy S. The pre-processing techniques for breast cancer detection in mammography images. *Int J Image Graph Signal Process.* 2013;5(5):47.
50. George MJ, Sankar SP. Efficient preprocessing filters and mass segmentation techniques for mammogram images. In: Circuits and Systems (ICCS), 2017 IEEE International Conference On. IEEE; 2017. p. 408–13. <https://doi.org/10.1109/iccs1.2017.8326032>.
51. Gonzalez RC, Woods RE, et al. Digital image processing, 2nd ed. Prentice hall Upper Saddle River; 2002.
52. Zuiderveld K. Contrast limited adaptive histogram equalization. 1994;474–85. <https://doi.org/10.1016/b978-0-12-336156-1.50061-6>.
53. Abdelhafiz D, Nabavi S, Ammar R, Yang C. The effect of pre-processing on breast cancer detection using convolutional neural networks. In: Poster session presented at the meeting of the IEEE International Symposium on Biomedical Imaging. Washington, DC: IEEE; 2018.
54. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep learning to improve breast cancer detection on screening mammography. *Sci Rep.* 2019;9(1):1–12.
55. Jung H, Kim B, Lee I, Yoo M, Lee J, Ham S, Woo O, Kang J. Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. *PLoS ONE.* 2018;13(9):.
56. Carneiro G, Nascimento J, Bradley AP. Unregistered multiview mammogram analysis with pre-trained deep learning models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2015. p. 652–60. [https://doi.org/10.1007/978-3-319-24574-4\\_78](https://doi.org/10.1007/978-3-319-24574-4_78).
57. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint. 2014. arXiv:1409.1556.
58. Guerrero-Pena FA, Fernandez PDM, Ren TI, Yui M, Rothenberg E, Cunha A. Multiclass weighted loss for instance segmentation of cluttered cells. arXiv preprint. 2018. arXiv:1802.07465.
59. Kinga D, Adam JB. A method for stochastic optimization. In: International Conference on Learning Representations (ICLR), vol. 5; 2015.

60. Csurka G, Larlus D, Perronnin F, Meylan F. What is a good evaluation measure for semantic segmentation? In: *BMC*, vol. 27. Citeseer; 2013. p. 2013. <https://doi.org/10.5244/c.27.32>.
61. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint*. 2017. arXiv:1704.06857.
62. Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, Blaschko MB. Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2019. p. 92–100. [https://doi.org/10.1007/978-3-030-32245-8\\_11](https://doi.org/10.1007/978-3-030-32245-8_11).
63. Geras KJ, Wolfson S, Kim S, Moy L, Cho K. High-resolution breast cancer screening with multi-view deep convolutional neural networks. *arXiv preprint*. 2017. arXiv:1703.07047.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

