

DU-Net: Convolutional Network for the Detection of Arterial Calcifications in Mammograms

Manal AlGhamdi^{ID}, *Member, IEEE*, Mohamed Abdel-Mottaleb, *Fellow, IEEE*, and Fernando Collado-Mesa

Abstract—Breast arterial calcifications (BACs) are part of several benign findings present on some mammograms. Previous studies have indicated that BAC may provide evidence of general atherosclerotic vascular disease, and potentially be a useful marker of cardiovascular disease (CVD). Currently, there is no technique in use for the automatic detection of BAC in mammograms. Since a majority of women over the age of 40 already undergo breast cancer screening with mammography, detecting BAC may offer a method to screen women for CVD in a way that is effective, efficient, and broad reaching, at no additional cost or radiation. In this paper, we present a deep learning approach for detecting BACs in mammograms. Inspired by the promising results achieved using the U-Net model in many biomedical segmentation problems and the DenseNet in semantic segmentation, we extend the U-Net model with dense connectivity to automatically detect BACs in mammograms. The presented model helps to facilitate the reuse of computation and improve the flow of gradients, leading to better accuracy and easier training of the model. We evaluate the performance using a set of full-field digital mammograms collected and prepared for this task from a publicly available dataset. Experimental results demonstrate that the presented model outperforms human experts as well as the other related deep learning models. This confirms the effectiveness of our model in the BACs detection task, which is a promising step in providing a cost-effective risk assessment tool for CVD.

Index Terms—Cardiovascular, deep learning, mammogram, segmentation. U-Net.

I. INTRODUCTION

MAMMOGRAPHY is an X-ray imaging technique used to examine the human breast for diagnosis and screening. Calcifications noted on mammograms may be defined as either suspicious (for breast carcinoma) or benign. Breast

arterial calcifications (BACs) are among the benign calcifications, which deposit linearly along the wall of breast arteries [1]. The BACs phenomenon has been known to radiologists for many years; however, its clinical value continued to be undefined until the first clinical study of the significance of BACs was performed in 1980 by Baum *et al.* [2]. This study revealed a strong association between BACs and diabetes mellitus. Subsequently, various studies have concluded that the presence of BACs on mammograms can be a useful indicator of coronary heart disease [3]–[7]. Most of these studies showed that after multivariate analysis, only BAC was found to have a significant association with both CAC (OR 2.87, 95% CI, [1.67 – 4.94], $P < 0.001$) and coronary artery plaque (OR 2.52, 95% CI, [1.53 – 4.18], $P < 0.001$). They also demonstrated that combining the presence of BACs to the 10-year atherosclerotic cardiovascular disease (ASCVD) risk significantly increases the areas under the curve leading to improvement in the net reclassification index [4].

Recently, several studies investigated the possibility of automatic BAC detection. Ge *et al.* [8], for instance, combined a set of image filtering methods with a k-segments clustering algorithm followed by a classifier to detect BACs. Cheng *et al.* [9]–[11] presented a two-step method that formulates the task as vessel tracking using the calcification points. In the first step, they use a tracking algorithm with an uncertainty scheme to detect the possible BACs paths. Then, the second step they use a compiling and linking algorithm to connect the generated paths into BACs. Finally, Mordang *et al.* [12] detect BACs as false candidates during microcalcification detection in breast cancer screening by applying a GentleBoost classifier to a set of pre-extracted features. Despite the effectiveness of these studies, the automatic detection of BACs in mammograms is still a challenging task. This is due to how BACs appear in mammograms in terms of their size, shape and contrast variations. Fig. 1(b)–(e) show magnified segments from one of the standard mammographic views known as the mediolateral oblique (MLO). The BACs show as linear deposits of bright spots along the walls of the arteries in the breast. However, these linear deposits have different variations and may appear as broken lines (Fig. 1(b)), connected and strong lines (Fig. 1(c)) or thin and weak lines (Fig. 1(d)), or lines which intersect with one another (Fig. 1(e)).

Convolutional neural network (CNN) models have been successfully employed in the interpretation of medical images

Manuscript received January 15, 2020; revised March 12, 2020 and April 16, 2020; accepted April 16, 2020. Date of publication April 23, 2020; date of current version September 30, 2020. (Corresponding author: Manal AlGhamdi.)

Manal AlGhamdi is with Department of Computer Science, Umm Al-Qura University, Makkah 21421, Saudi Arabia (e-mail: maalghamdi@uqu.edu.sa).

Mohamed Abdel-Mottaleb is with the Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146 USA (e-mail: mottaleb@miami.edu).

Fernando Collado-Mesa is with the Department of Radiology, Miller School of Medicine, University of Miami, Coral Gables, FL 33146 USA (e-mail: fcollado@med.miami.edu).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2020.2989737

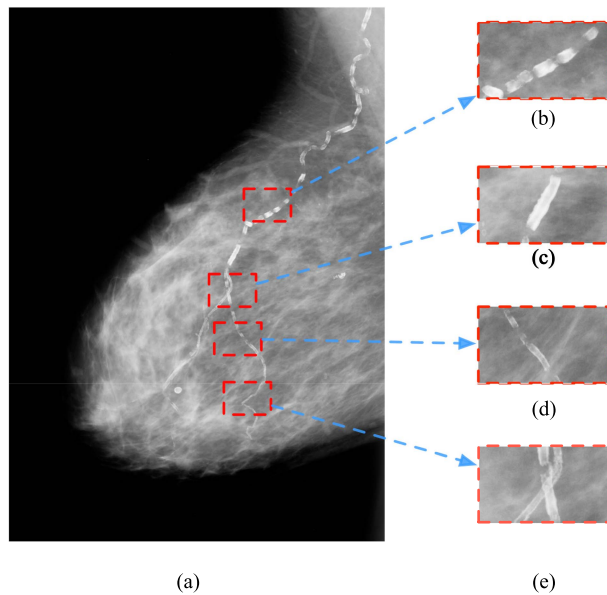


Fig. 1. An illustration of BACs from a clipped mammogram MLO view. (a) A mammogram. (b)-(e) Examples of different appearance patterns of calcified vessels [9].

such as head computerized tomography scans [13], fundus images [14] and mammograms [15]. These models have the ability to learn the deep features of the input without any pre-processing steps such as segmentation or feature extraction. However, constructing these models to achieve the desired results is restricted by the availability of large annotated training sets which are not always available. To address this problem, Ronneberger *et al.* [16] extended the state-of-the-art model in semantic segmentation known as the 'fully convolutional network' (FCN) [17] to learn the model with an efficient generalization performance using a limited number of labeled images. This model is known as 'U-Net' and it contains an encoder, or 'contracting' path, to extract the features from the whole image and a decoder, or 'expanding' path, to generate a segmentation that has the input resolution. Unlike the traditional encoder-decoder networks, the U-Net applies what is known as long-skip connections to directly connect opposing layers from the contracting to the expanding paths [18]. Therefore, the spatial information lost during the sub-sampling at the contracting path is recovered in the expanding path by passing equal resolution features from the former to the latter. This allows the U-Net and its variants (MDU-Net [19], 3D-U-Net [20], [21] and U-Net++ [22]) to achieve remarkable results for many biomedical segmentation tasks, including blood vessel segmentation in retina images, skin cancer segmentation, and lung lesion segmentation [23].

Deeper networks, such as the U-Net, have been shown to be more beneficial for image segmentation and semantic segmentation [24]–[26]. However, these networks suffer from the problem of vanishing gradients where the gradients of the loss function approach zero, making the network hard to train [27]. To address this problem, short-skip connections were introduced in ResNet [26] containing residual blocks that sum two signals: a non-linear transformation of the input

and its identity mapping that is implemented by shortcut connection. Then, the DenseNet [28] was presented as an extension of the ResNet by replacing the summation of the previous feature maps with concatenation. This small change allow the DenseNet to enhance feature propagation, strengthen feature reuse and to be a good fit for different biomedical segmentation tasks [29]. For instance, Li *et al.* [30] proposed H-DenseUNet for liver and tumor segmentation by extending only the contracting path of the U-Net with dense connection as 2D DenseUNet to extract the intra-slice features followed by 3D DenseUNet to hierarchically aggregate the volumetric contexts. Guan *et al.* [31] combined the DenseNet and the U-Net as a post-processing step for removing artifacts from images. Meanwhile, Chen *et al.* [32] presented a biomedical segmentation method by extending the U-Net using a dense connection block at the contracting path and a residual connection block in the expanding path.

Deep learning models have achieved promising results in various biomedical tasks, however, BAC detection is still a very challenging task. To the best of our knowledge, there is only one attempt by Wang *et al.* [15] to generalize a CNN architecture for the BACs detection. In their work [15], a pixelwise, patch-based, two-class classification CNN model was developed with twelve layers to identify the presence of BACs in mammograms. This means that each pixel is classified by processing the image patch around it. Their results show that the accuracy of their DL model is as good as that produced by human experts. However, it only focuses on the neighborhood of the specified pixel, depending on the patch size, and thus limits the data used by the network for the global representation of the input.

In this paper, we explore the applicability of the U-Net to automatically detect BACs in mammograms in support of the future goal to develop an automated BAC detector as a risk marker for CAD [1], [33], [34]. Inspired by the success of the U-Net-based and DenseNet-based models in many biomedical segmentation tasks, we present DU-Net that combines both methods with their short- and long-skip connections. Dense connections (or short-skip connections) prevents the model from learning redundant features and enhances information flow leading to more efficient CNN. Different from the conventional U-Net model, we use the summation operation instead of the concatenation at the end of the long-skip connections that combines the same-level feature maps. This leads to better performance and reduces the computational load. To evaluate the performance of the presented model, we conducted a reader study by human expert radiologists to create the ground-truth from a publicly available dataset. We evaluated the performance on a dataset consisting of 826 full-field digital mammograms using different types of analysis methods. Experimental results demonstrate that the presented model outperforms human experts with more than three years of experience as well as the other related deep learning models. Following is a list of the contributions of this paper:

- We developed a method for detecting BACs, which can be used in the future as a risk marker for CAD.

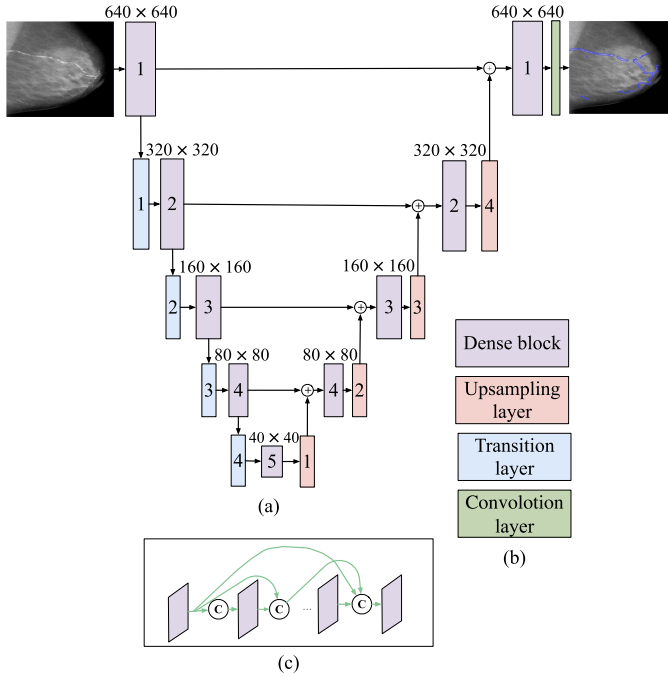


Fig. 2. A Flow chart of (a) the proposed DU-Net model that extended the U-Net [16] model with dense blocks [28] and summation long-connections, (c) the detailed structure of the Dense block where 'C' means concatenation operation.

- We present DU-Net that extends the U-Net using the dense blocks and two types of connections, *i.e.*, concatenation is used to combine information within the dense blocks and summation is used to combine information shared between the two paths.
- From a publicly available dataset for breast cancer detection, we created a BACs ground-truth dataset annotated by radiologists. This dataset along with the annotations will be made available to other researchers.
- Compared with human performance, the presented model confirms the deep learning ability to perform better than human experts for this task.

The remainder of the paper is organized as follows: the proposed DU-Net model is introduced in Section II. Datasets and the experimental setup are presented in Section III, followed by results and discussion in Section IV. Finally Section VI concludes the paper.

II. APPROACH

In this work, the BAC detection task is redefined as a semantic segmentation task, where each pixel is classified as one of multiple object classes. The BAC detection task has two classes: BAC and non-BAC. Based on the DenseNet [28] and the U-Net [16], we have proposed a DU-Net model for BAC detection. Details of our approach are illustrated in Fig. 2 and the following subsections.

A. The DU-Net Model

We extend the conventional U-Net [16] by developing a contracting path of dense blocks [28] with spatial reduction to

extract the features, followed by a corresponding expanding path with dense blocks to generate an output image with the same size as the input image [16]. Both paths contain a matching number of dense blocks which in turn contain a matching number of layers. As confirmed in [16], [18], the information lost through the contracting path is recovered in the expanding path by sending feature maps with the same resolution from the contracting path to the expanding path. Thus, the feature maps computed at the expanding path can be considered as refinements of the ones computed at the contracting path [35]. Existing literature indicates that summation operations between the features lead to better performance compared with concatenation, and also significantly reduce the computational load [36], [37]. Therefore, for the skip connections between the contracting path and the expanding path in our DU-Net, the concatenation operation between the same-level feature maps of the U-Net is replaced with summation operation. Hence, the local details found in the feature maps of the contracting path are directly inserted in the feature maps of the expanding path.

In the conventional CNN model, the output of the l^{th} layer is computed by a non-linear activation function as:

$$x_l = H_l(x_{l-1}) \quad (1)$$

where x_l is the output of l^{th} layer, x_{l-1} is the output of $(l-1)^{th}$ layer, and H represents a convolution followed by rectified linear unit (ReLU). The DenseNet [28] contains dense blocks that use all the preceding feature maps in its output as

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (2)$$

where $[\dots]$ is defined as the concatenation operation and H_l represents a function of three sequential operations: batch normalization (BN), followed by a ReLU and a 3×3 convolution. Each dense block produces k feature-maps where k is known as the growth rate of the network. Given the number of channels in the input layer as k_0 , then the number of the input feature maps of the l^{th} layer will be:

$$k_0 + k \times (l - 1). \quad (3)$$

In this work, we adopted the general structure of the DenseNet-121 [28], *i.e.*, as the contracting and expanding paths after removing the classification layer, leaving only the dense blocks and transition layers. The detailed structure of the DU-Net is presented in Table I.

In the contracting path, each dense block has repeated convolutional blocks consisting of batch normalization (BN) + ReLU + 1×1 conv + BN + ReLU + 3×3 conv with a growth rate of $k = 32$. The growth rate is updated at each layer so that all dense blocks have the same number of convolutional layers. There are 6, 12, 24 and 16 convolutional blocks in four dense blocks, respectively. Within the dense block, the input feature map feeds into the sequence of operations mentioned above, which produces the output feature maps. A transition down layer is used after every dense block to reduce the size of the feature map. Thus, the dimension of the input is reduced by half and the number of the feature maps is doubled. Each transition layer consists of BN + ReLU + 1×1 conv + 2×2 max pooling layers.

TABLE I

DU-NET ARCHITECTURE, THE GROWTH RATE IS SET TO $K = 32$.
NOTE THAT EACH CONV LAYER SHOWN IN THE TABLE
REPRESENTS BN + RELU + CONV

Layers	Ingredients	Output size
input		640×640
dense block 1	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	640×640
transition layer 1	$1 \times 1 \text{ conv}$	640×640
	$3 \times 3 \text{ max pooling}$	320×320
dense block 2	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	320×320
transition layer 2	$1 \times 1 \text{ conv}$	320×320
	$3 \times 3 \text{ max pooling}$	160×160
dense block 3	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	160×160
transition layer 3	$1 \times 1 \text{ conv}$	160×160
	$3 \times 3 \text{ max pooling}$	80×80
dense block 4	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	80×80
transition layer 4	$1 \times 1 \text{ conv}$	80×80
	$3 \times 3 \text{ max pooling}$	40×40
dense block 5	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	40×40
dense block 4	$3 \times 3 \text{ upsampling 1 - [dense block 4]}$	80×80
dense block 3	$3 \times 3 \text{ upsampling 2 - [dense block 3]}$	160×160
dense block 2	$3 \times 3 \text{ upsampling 3 - [dense block 2]}$	320×320
dense block 1	$3 \times 3 \text{ upsampling 4 - [dense block 1]}$	640×640
output	$1 \times 1 \text{ conv}$	640×640

After passing through the contracting path, the dense block 5 (*i.e.*, 40×40) is considered as the bridge block between the two paths, which starts expanding the feature maps in the following expanding path. In the expanding path, 3×3 upsampling layers are used before each dense block to expand the size of the image. Similar to the contracting path, each dense block in the expanding path contains BN + ReLU + 1×1 conv + BN + ReLU + 3×3 conv with 16, 24, 12 and 6 growth rates, respectively. Each block computes the summation of the outputs of the previous transition layer and the same-level dense block in the contracting path. Before the summation operation, the number of the feature maps is reduced to half using 1×1 convolution to maintain the symmetry of the network. The final 1×1 convolution layer is used to map the feature vector to the binary prediction (*i.e.*, BAC vs. non-BAC).

B. Loss Function

One of the main challenges that we faced in this work is the extreme class imbalance between the positive BAC pixels and the negative background pixels. In the BAC detection task, the loss is computed from both background and BACs. On average, the relevant pixels corresponding to BACs only comprise 2% of the total pixels. Different loss functions have been proposed to address this problem. Focal loss (FL) [38] was introduced as an extension of the cross entropy (CE) loss function to address the imbalance between the classes that represent the foreground and background during the training.

Given the ground-truth class $y \in \{\pm 1\}$ and the estimated probability of the binary classification $p \in [0, 1]$, p_t [38] is defined as follows:

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (4)$$

The CE loss function is computed as:

$$CE(p_t) = -\log(p_t) \quad (5)$$

The FL introduces a modulating factor $\alpha_t(1 - p_t)^\gamma$ into the CE loss function as follows:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (6)$$

where $\alpha_t \in [0, 1]$ is a weighting factor for class 1, and $(1 - \alpha_t)$ is the weighting factor for class -1 , and $\gamma \geq 0$ is a tunable parameter. In practice α_t can be set by inverse class frequency. This modulating factor reduces the error contribution from the background samples and the easily classified samples to prevent them from overwhelming the loss function, thus, it increases the model's focus on correcting the misclassified samples during training.

III. EXPERIMENTS

To evaluate the BAC detection performance of our proposed models, we conducted experiments using a subset of a publicly available dataset that we prepared for this task.

A. Dataset

The Digital Database for Screening Mammography (DDSM) [39] is a well-known publicly available dataset for breast disease detection. In this paper, we use a later version of this dataset known as Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [40] that consists of images which were converted into the standard DICOM format. The scans were collected from 1,700 cases with three conditions relating to breast cancer (normal, benign, and malignant). Most of the cases include the cranial-caudal (CC) and mediolateral oblique (MLO) views of each breast, as well as the patient and the image information. Because of the sample size limit, we processed each view as a separate image. The dataset also includes ROI segmentation and bounding boxes and pathologic diagnosis for 1,520 images that represent the ground-truth training samples. This dataset has relatively large-sized images with a mean height of 5295 pixels and a mean width of 3131 pixels. The purpose of our study is to detect BACs in mammograms as a risk assessment for CAD; thus, we do not use the available ground-truth.

B. Ground-Truth

To train and evaluate the BAC detection algorithm, we performed a reader study on the CBIS-DDSM images. Similar to [15], three radiologists with expertise in reading mammograms (Department of Radiology, University of Miami Miller School of Medicine, Miami, FL, USA) were involved in preparing the ground-truth for this study. We will refer to these readers as A, B, and C; reader A has about ten years of experience in reading mammograms, while readers B and C have about three years of experience. The readers manually draw the BACs boundaries in the mammogram images using the labelbox tool [41], as shown in Fig. 3.

Considering the experience of the readers and the subtlety of the BACs, we followed a three-step process:

- Since the CBIS-DDSM dataset was not originally collected for the BACs detection task, the first step in generating the ground-truth was to identify the images

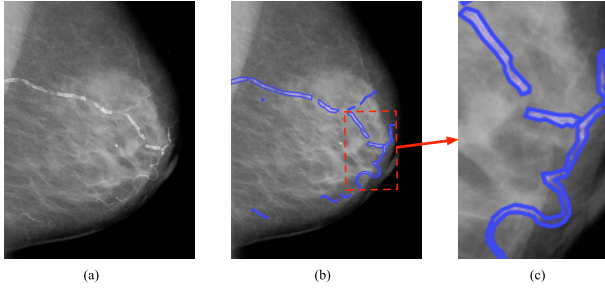


Fig. 3. An illustration of BACs ground-truth from a mammogram MLO view prepared by human experts. (a) A mammogram image. (b) Full annotated image. (c) Magnified part for the annotation.

that contain BACs from the entire dataset. For this step, reader A was asked to go through all 3,400 images and choose the ones that contain BACs, which yielded only 413 images.

- In the second step, these 413 images were randomly divided into two non-overlapping subsets with 206 and 207 images. One subset was assigned to reader B and the other subset was assigned to reader C. Each reader was asked to examine the images and draw an accurate boundary around each BAC.
- In the third step, the annotations were evaluated and corrected by reader A, resulting in a second reading for each mammogram that was considered to be the ground-truth for both training and testing in this study.

At the end of this process, the generated dataset contained 413 mammogram images with BACs, and the same number of images without BACs were chosen randomly from the remaining images in the dataset to generate a BAC detection dataset of 826 mammogram images.

C. Evaluation Criteria

The proposed DU-Net was evaluated using sensitivity, specificity, overall accuracy, the Dice similarity coefficient (or F1-score), the Jaccard index, and the Matthews correlation coefficient (MCC) [42]. These measurements are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (7)$$

$$\text{Sensitivity (or TPR)} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (9)$$

$$\text{F1-score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (10)$$

$$\text{Jaccard index} = \frac{TP}{TP + FP + FN} \quad (11)$$

$$\text{TNR} = \frac{TN}{TN + FP} \quad (12)$$

$$\text{FNR} = \frac{FN}{FN + TP} \quad (13)$$

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (14)$$

where TP is the true positive, TN is the true negative, FP is the false positive and FN is the false negative. These values are calculated in terms of number of pixels. First, the detected BAC area is considered to be correctly detected if 90% of the pixels detected by the model overlap with the BAC ground-truth; in that case the pixels in the detected area that overlap with the ground truth area are counted as TP. If the detected area has less than 90% overlap with the ground truth area, all the pixels in the detected area are considered to be FP. On the other hand, pixels in the ground truth area that are not detected are counted as FN, and pixels that are in the background and are not detected as BAC are counted as TN. A good performance is achieved when obtaining high sensitivity and specificity. Note that, the sensitivity is equivalent to the true positive rate (TPR) and 1-specificity is equivalent to the false positive rate (FPR).

The overlapping between the predicted and ground-truth areas is measured using the Dice (F1- score) and Jaccard metrics, which compute the ratio between the TP pixels and the total number of detected pixels [43]. The MCC measures the correlation coefficient between ground-truth and the predictions and output a value between -1 and $+1$, where the higher values indicate better correlation. In addition, the free-response receiver operating characteristic (FROC) analysis [44], which is an extension of the receiver operating characteristic (ROC), is also employed to measure the overall performance. The ROC is used for diagnosis studies, while the FROC is used for detection studies. In an ROC curve, sensitivity is plotted on the y-axis and 1-Specificity (FPR) is plotted along the x-axis, while the FROC curve replaced the FPR by the number of false positives per image. The inverse ROC (or ROC^{-1}) curve, which plots the TNR (where $FPR=1-TNR$) as a function of the FNR (where $FNR=1-TPR$), is also used to measure the performance of the model using images with no BACs.

D. Experiment Set-up

Training and deployment of the network was performed using a PC equipped with an NVIDIA GTX 1080TI 11GB GPU card with a 3584 CUDA parallel processing core. We implemented the model using the Keras open-source deep learning library [45], which is an easy high-level programming API built on Python with an option of Theano or TensorFlow as a backend deep learning engine. To measure the effectiveness of our model, we used the K-folds cross-validation procedure where the entire dataset was randomly divided into equally sized k-folds, where $k = 5$. The model was trained using (k-1) folds and tested using the remaining kth fold. The procedure was repeated until all folds were used as a testing set and the performance was computed as the average over all folds. The training was performed using the stochastic gradient descent [46] with a batch size of 256 and high momentum (0.9). The initial learning rate (lr) was 0.01 and was decayed according to the equation:

$$lr = lr * (1 - \text{iterations}/\text{total_iterations})^{0.9} \quad (15)$$

For the FL in Eq. 6, the α was set to 0.25 and the γ was set to 2, which were the best values obtained in the original paper [38].

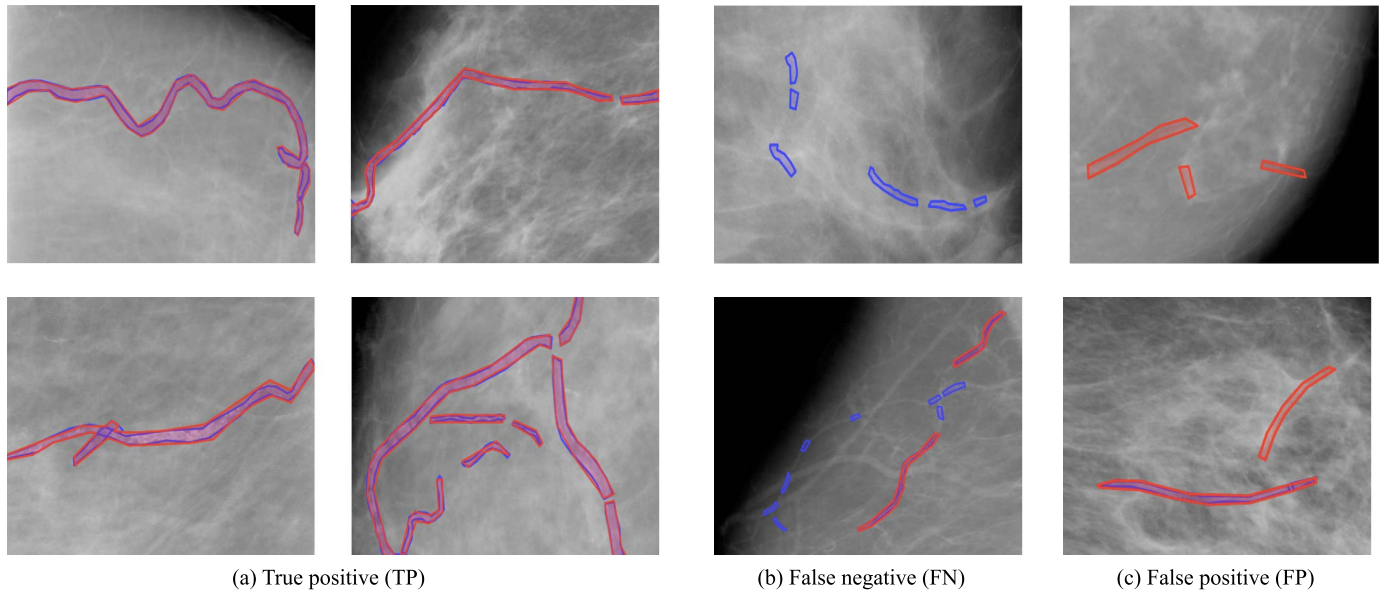


Fig. 4. Illustration of the detection results of our DU-Net model using mammogram images. The ground-truth BACs are contoured by blue lines, while the automatically detected BACs are contoured by red lines. The magnified images show (a) results with correctly detected BACs, (b) results with some mis-detected BACs, and (c) results with falsely detected BACs (seen better in color).

Data augmentation was used to overcome the overfitting by artificially increasing the training samples with class-preserving transformations. This helps ‘produce’ more training data, prevent overfitting and account for obvious invariances for the classification task at hand. In our experiment, data augmentation was used by applying random vertical shift, horizontal shift, horizontal flip, and zoom-in. These affine transformations help the model to have a better understanding of the input image since it views the images in many transformed views including scaling. The range between $[-20, 20]$ was used for vertical and horizontal shift and the range between $[0, 0.5]$ was used for the zoom in.

E. Comparison

For further quantitative evaluation of the DU-Net’s performance, we performed three different comparisons. The first comparison was with Wang *et al.* [15], which is the only related study to apply deep learning for this task. Comparisons with their published results are not feasible because they used their own private dataset. Thus, we followed their published paper [15] to implement their CNN model and evaluate it using our dataset.

The second comparison was performed to evaluate the effectiveness of each component in the DU-Net model, *i.e.*, the summation operation at the long-skip connections and dense short-skip connections, by constructing the following models:

- **U-Net [16]**, which is the original model containing long-skip concatenation connections and without dense short-skip connections.
- **U-Net [16] + summation operation (or U-Net+S)** that incorporates the summation operation at the long-skip connections.

- **U-Net [16] + dense blocks (or U-Net+D)** that incorporates dense short-skip connections.

The same set-up described in Section III-D was used for all these models.

The third comparison was performed against the annotations generated by readers B and C, before being approved by reader A as described in Section III-B. This is because reader A has more experience in BACs detection than reader B and C and thus was responsible for creating the gold standard annotations.

IV. RESULTS

Examples of BACs detection results using the DU-Net model on mammogram images are presented in Fig. 4. The blue contour in the figure is drawn by the human expert as ground-truth and the red contour is automatically detected by the proposed model. The first two columns show examples where the DU-Net’s performance is comparable with the human experts’ performance in correctly detecting the BACs. These examples show accurate BACs detection results at the exact positions of the ground truth for each mammogram. The third column shows some mis-detected BACs results which were considered by the DU-Net model as background. This may be caused by cases which contain small groups of BACs with non-continuous structures. The last column shows microcalcifications which were misidentified as BACs in some mammograms. Mammograms of patients with dense breast tissues are possibly having this type of error. With sufficient number of such cases for training this error can be avoided.

A. FROC Analysis

Fig. 5 presents the FROC curves obtained from the 826 mammogram images, defined in Section III-B, using

TABLE II

COMPARISON OF THE PROPOSED DU-NET PERFORMANCE (IN %) WITH RELATED METHODS ON OUR ANNOTATED DATASET. ALL RESULTS ARE REPORTED FROM THE AVERAGE OF THE 5-FOLDS CROSS-VALIDATION

Method	Accuracy	Sensitivity	Specificity	F1-score	Jaccard index	MCC
CNN [15]	62.61	57.24	66.7	56.8	62.2	60.95
U-Net [16]	72.1	75.33	70.41	71.87	67.25	65.76
U-Net + summation long-connections (or U-Net + S)	72.8	76.13	71.53	72.1	68.5	63.25
U-Net + dense short-connections (or U-Net + D)	80.17	82.08	83.23	79.3	75.89	72.71
our DU-Net	91.47	91.22	92.01	92.19	85.17	84.93

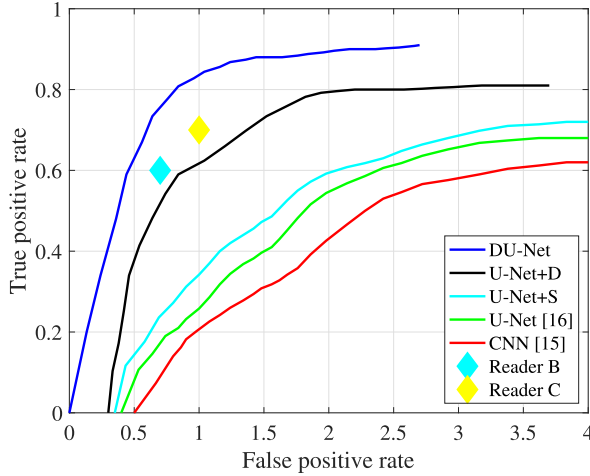


Fig. 5. Evaluating the DU-Net model's performance for BAC detection in terms of FROC curves with the false positive rate (FPR) on the X-axis and the true positive rate (TPR) on the Y-axis.

our DU-Net, the CNN [15], the three comparison models and the ground-truth annotators (B and C in Section III-B). Comparing these curves, the best performance was achieved by the DU-Net model which combines the advantages of both dense connections and U-Net paths. The dense connectivity mitigates the problem of overfitting when using a small dataset and encourages the model to learn a more discriminative set of features. Meanwhile, the summation long-connections improve information flow and promote feature reuse throughout the network by adding the local details found in the contracting path to the expanding path. As shown, our DU-Net model achieved FPR=2 per image at a 90% TPR detection rate and FPR=0.7 per image at an 80% TPR detection rate. In comparison, at the same FPR=2 per image, the CNN [15] achieved 42% TPR, U-Net achieved 56% TPR, U-Net+S achieved 59% TPR and U-Net+D achieved 80% TPR.

Compared with the human experts, the DU-Net outperforms readers B and C. Please note that there are no FROC curves for readers B and C, and their performances are represented in the FROC figure using a single operational point for each. As illustrated in Fig. 5, reader B achieves TPR=60.21% with FPR=0.7, which is lower than the DU-Net with TPR=77.8% at the same FPR. Meanwhile, reader C achieves TPR=70.1% with FPR=1, which is lower than that achieved by the DU-Net with TPR=84.4% at the same FPR rate. Considering this result and the tedious process used by experts in a clinic to detect BAC, the DU-Net model is more efficient in detecting and locating the BACs. This is an extremely promising outcome

with regard to deploying such a model in clinical environments and using it to assess the risk of CAD.

B. Other Analysis

The performances obtained from the 826 mammogram images, defined in Section III-B, in terms of sensitivity, specificity, accuracy, F1-score, and Jaccard values are reported (in %) in Table II. As shown, the DU-Net clearly outperforms the other methods (defined in Section III-E) with an overall accuracy of 91.47%, sensitivity of 91.22%, specificity of 92.01%, F-1 score of 92.19, Jaccard coefficient of 85.17%, and MCC of 84.93%. These results confirm that the DU-Net model is more efficient than the CNN for the BACs detection task. Also, they show the significant improvements achieved over the U-Net by adding the summation long connections and the dense short connections. The results are also supported by the improvement in the MCC, clarifying that our detection results attained very good matching with the ground-truth.

It is also important to measure the performance of the DU-Net model on mammograms without BACs to evaluate the model's robustness against the non-BACs features including masses and microcalcifications. For this purpose, the DU-Net model was tested on mammograms without BACs (413 images), where ideally the model should label all the pixels as background. For that, the negative evaluation criteria defined in Section III-C was used. The TNR represents the number of the pixels correctly detected by the model as background pixels, while the FNR reflects the number of the pixels that represent masses or microcalcification but incorrectly detected by the model as BACs [47], [48]. The results are presented in Fig. 6 where the best performance would have high TNR and low FNR. Compared with U-Net [16] and the CNN in [15], our DU-Net model achieved FNR=2 per image at a 85% TNR and FNR=0.7 per image at an 75% TNR. In comparison, at the same FNR=2 per image, the CNN [15] achieved 40% TNR, U-Net achieved 55% TNR. This shows that the DU-Net model is robust when applied to BAC and non-BAC cases.

C. Feature Learning

To demonstrate the ability of the DU-Net model to learn discriminative features, Fig. 7 shows some examples of feature maps generated after the max-pooling layers in the contracting path, which is deep enough to generate and visualize distinctive features used for classification. Fig. 7(a) shows examples of features learned from BACs samples, while Fig. 7(b) show

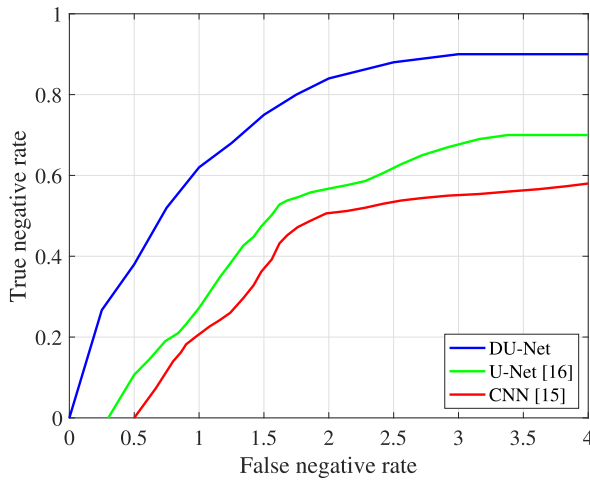


Fig. 6. Evaluating the DU-Net model's performance for non-BAC detection with the false negative rate (FNR) on the X-axis and the true negative rate (TNR) on the Y-axis.

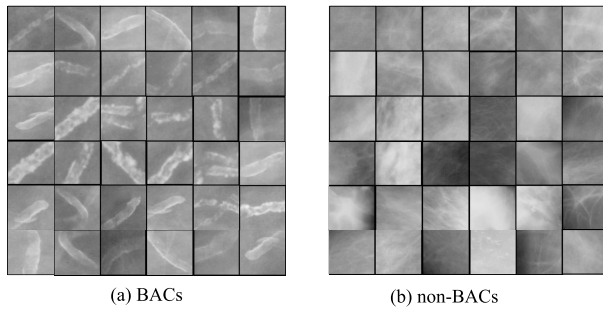


Fig. 7. Examples of features learned by the DU-Net model from (a) BAC samples, and (b) non-BAC samples.

TABLE III

COMPARISON OF SEGMENTATION TIME FOR THE DU-NET AGAINST OTHER METHODS

Method	Training time per epoch (Second)	Testing time per epoch (Second)
CNN [15]	580	14
U-Net [16]	471	10.62
U-Net + S	310	10.20
U-Net + D	494	10.9
our DU-Net	482	10.81

examples of features learned from non-BAC samples. Comparing the two feature maps, it is clear that the DU-Net model learns the BACs patterns with different scale, orientation and viewpoint variations. From the non-BACs features, the model was able to recognize the absence of BACs and distinguish them from the other types of calcifications that also appear in the mammograms.

D. Training and Testing Time

Table III presents a comparison of the training time and testing time for the DU-Net and the four related models (defined in Section III-E) for 100 epochs. As the table shows, the slowest model was the CNN [15] while the fastest one was the U-Net+S. Our DU-Net model achieves good BACs detection results in a reasonable training and testing time

and is faster than both CNN [15] and U-Net+D. The U-Net, which was originally developed as a fast segmentation model for medical image processing, is the second fastest model after the U-Net+S. The concatenation operation in the U-Net requires longer computation time to generate feature maps than the summation operation in the U-Net+S. Adding the dense connectivity increases the computation time in the U-Net+D but the DU-Net overcomes this problem by combining both dense connectivity and the summation long-connections.

V. DISCUSSION

The BAC phenomenon has attracted the attention of radiologists and other medical specialists to investigate their relationship to CVD. Indeed, various medical studies have concluded that the presence of BACs on mammograms can be a useful indicator of CVD and CAD. Despite this, the automatic detection of BACs in mammograms is still considered to be an unsolved and challenging research task. Using deep learning-based approaches, only one related work has been published by Wang *et al.* [15] where a pixel-wise and patch-based CNN model was developed with twelve layers to identify the presence of BACs mammograms. Inspired by the successful results achieved by the U-Net-based and DenseNet-based models in many biomedical segmentation tasks [30]–[32], we present DU-Net which combines both methods with their short-dense and long-summation connections. The dense connectivity was added to both the contracting and expanding paths of the U-Net CNN architecture, while the long-summation connections linked the contracting part and the expanding part to preserve local features. The dense connections help the DU-Net model to avoid learning redundant features and ensure maximum information flow between internal layers. This also improves the gradient flow and alleviates the effort required to search the optimal solution in a very deep neural network. Meanwhile, the long-summation connections help the local details found in the feature maps of the contracting path are directly inserted in the feature maps of the expanding path. This helps to localize and connect the small parts of the detected BACs.

Currently, there is no public dataset for BACs detection. Thus, we conducted a reader study, with the help of expert radiologists, to provide ground-truth information for a publicly available dataset. This information will be available to the research community and can be used by other researchers to contribute to this area. We compared our results with the method by Wang *et al.* [15], U-Net [16], U-Net with only summation long connections and finally U-Net with only dense short connections. This helps to evaluate the effectiveness of each extension we performed on the U-Net. We also evaluated the performances against two human expert radiologists.

The results presented in Section IV show that the DU-Net outperformed the related methods as well as the human experts. The lowest performance was achieved by the CNN model used in [15]. This is expected since the typical use of CNNs is to solve classification problems, where the model is used to classify each image into a single class label. Meanwhile, the desired output of our task should include

localization, where each pixel in the image should be classified with a class label. Since it is a patch-based model, there is a trade-off between the accuracy of locating the BAC and the use of context. Larger patches decrease the accuracy of localization since they need more max-pooling layers, while small patches give the model a little context only. Moreover, training CNN from scratch requires a large number of training images which is not possible in many medical applications. The U-Net achieved better results thanks to the long concatenation connections between the encoding and decoding parts, which enables the model to reuse the features for better detection and localization. The summation operation in the U-Net+S model retained and slightly improved the performance by replacing the long concatenation connections with summation connections. The summation operation helped the U-Net to speed up the training process and preserve the local details found in the feature maps between the contracting and the expanding paths. The best performance after our DU-Net model was achieved by the U-Net+D due to the dense connections that avoid learning redundant features and ensure maximum information flow between layers.

Evaluating the processing time for all models revealed that the DU-Net performed as fast as the original U-Net. Replacing the concatenation long-connections with summations helped the model to perform faster. Meanwhile, the CNN was found to be quite slow since it needs to be run for each patch separately.

Finally, the presented model was trained and tested using a small-sized and low-quality dataset. Extending the U-Net with different types of skip connections helps to address the shortage of ground-truth since each layer re-use the features from all previous ones. However, a major improvement in model's performance can be obtained by using a larger dataset for training with a big number of diverse cases. Therefore, we have recently secured IRB approval to collect and annotate a large dataset of de-identified mammograms known to have vascular calcifications from the University of Miami Health Care System (UHealth), Miami, FL, USA, and we also encourage the research community to work towards this direction.

VI. CONCLUSIONS

This paper presented the DU-Net as an automated and accurate detection model for BACs in mammograms which could potentially be used to effectively screen women for CVD risk at no additional cost or exposure to radiation. The model was developed based on promising results achieved by the U-Net model in many biomedical segmentation problems and the DenseNet in semantic segmentation tasks. Combining both models facilitates the reuse of computation through long- and short-skip connections and improves gradient flow. Experiments using a set of full-field digital mammograms collected and prepared for this task from a publicly available dataset demonstrate the DU-Net's ability to improve accuracy over the conventional and related models. The results also demonstrate that the presented model performed better than human experts. This study confirms the efficacy of extending the U-Net by dense connections when applied to the BACs detection task and is a promising step in providing a risk assessment tool to identify individuals with cardiovascular risks.

Currently, there are no available techniques for the automatic BACs detection using mammography. Considering that the mammograms are often available for breast cancer screening programs, utilizing them to also screen for CVD is very cost-effective. More studies are still required to address this task and with our publicly available annotations, the research community has the opportunity to develop and validate more approaches for BACs detection. Extending this type of research to also quantify BACs and to eventually add this information to the standard CVD risk assessment methods is part of our ongoing research. This could potentially help patients and providers to initiate discussion about lifestyle changes and to implement targeted medical therapies for prevention of CVD at an earlier stage. The reporting of BACs on mammograms may further motivate patients, by having concrete evidence of CAD, and improve adherence to medical therapy and lifestyle modifications.

ACKNOWLEDGMENT

The authors would like to thank Edilberto R. Alvarez and David To from the University of Miami/Jackson Memorial Hospital Diagnostic Radiology Residency Program, Miami, FL, USA for their contribution in annotating the dataset used in the experiments of this study.

REFERENCES

- [1] Q. M. Bui and L. B. Daniels, "A review of the role of breast arterial calcification for cardiovascular risk stratification in women," *Circulation*, vol. 139, no. 8, pp. 1094–1101, Feb. 2019.
- [2] J. K. Baum, C. H. Comstock, and L. Joseph, "Intramammary arterial calcifications associated with diabetes," *Radiology*, vol. 136, no. 1, pp. 61–62, Jul. 1980.
- [3] T. Chadashvili, D. Litmanovich, F. Hall, and P. J. Slanetz, "Do breast arterial calcifications on mammography predict elevated risk of coronary artery disease?" *Eur. J. Radiol.*, vol. 85, no. 6, pp. 1121–1124, Jun. 2016.
- [4] Y. E. Yoon *et al.*, "Prediction of subclinical coronary artery disease with breast arterial calcification and low bone mass in asymptomatic women," *JACC: Cardiovascular Imag.*, vol. 12, no. 7, pp. 1202–1211, Jul. 2019.
- [5] D. Newallo *et al.*, "Mammographic detection of breast arterial calcification as an independent predictor of coronary atherosclerotic disease in a single ethnic cohort of African American women," *Atherosclerosis*, vol. 242, no. 1, pp. 218–221, Sep. 2015.
- [6] L. Margolies *et al.*, "Digital mammography and screening for coronary artery disease," *JACC: Cardiovascular Imag.*, vol. 9, no. 4, pp. 350–360, 2016.
- [7] A. H. E. M. Maas *et al.*, "Breast arterial calcifications are correlated with subsequent development of coronary artery calcifications, but their aetiology is predominantly different," *Eur. J. Radiol.*, vol. 63, no. 3, pp. 396–400, 2007.
- [8] J. Ge *et al.*, "Automated detection of breast vascular calcification on full-field digital mammograms," *Proc SPIE*, vol. 6915, Mar. 2008, Art. no. 691517.
- [9] J.-Z. Cheng, E. B. Cole, E. D. Pisano, and D. Shen, "Detection of arterial calcification in mammograms by random walks," in *Information Processing in Medical Imaging*, J. L. Prince, D. L. Pham, and K. J. Myers, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 713–724.
- [10] J.-Z. Cheng, C.-M. Chen, and D. Shen, "Identification of breast vascular calcium deposition in digital mammography by linear structure analysis," in *Proc. 9th IEEE Int. Symp. Biomed. Imag. (ISBI)*, May 2012, pp. 126–129.
- [11] J.-Z. Cheng, C.-M. Chen, E. B. Cole, E. D. Pisano, and D. Shen, "Automated delineation of calcified vessels in mammography by tracking with uncertainty and graphical linking techniques," *IEEE Trans. Med. Imag.*, vol. 31, no. 11, pp. 2143–2155, Nov. 2012.
- [12] J.-J. Mordang, A. Gubern-Mérida, G. den Heeten, and N. Karssemeijer, "Reducing false positives of microcalcification detection systems by removal of breast arterial calcifications," *Med. Phys.*, vol. 43, no. 4, pp. 1676–1687, Mar. 2016.

- [13] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.
- [14] M. Al Ghamdi, M. Li, M. Abdel-Mottaleb, and M. A. Shousha, "Semi-supervised transfer learning for convolutional neural networks for glaucoma detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 3812–3816.
- [15] J. Wang *et al.*, "Detecting cardiovascular disease from mammo-grams with deep learning," *IEEE Trans. Med. Imag.*, vol. 36, no. 5, pp. 1172–1181, May 2017.
- [16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
- [18] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," 2016, *arXiv:1608.04117*. [Online]. Available: <http://arxiv.org/abs/1608.04117>
- [19] J. Zhang, Y. Jin, J. Xu, X. Xu, and Y. Zhang, "MDU-net: Multi-scale densely connected U-net for biomedical image segmentation," 2018, *arXiv:1812.00352*. [Online]. Available: <http://arxiv.org/abs/1812.00352>
- [20] G. Zeng, X. Yang, J. Li, L. Yu, P.-A. Heng, and G. Zheng, "3D U-Net with multi-level deep supervision: Fully automatic segmentation of proximal femur in 3D MR images," in *Machine Learning in Medical Imaging*. Cham, Switzerland: Springer, Sep. 2017, pp. 274–282.
- [21] S. R. Hashemi, S. S. Mohseni Salehi, D. Erdogmus, S. P. Prabhu, S. K. Warfield, and A. Gholipour, "Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection," *IEEE Access*, vol. 7, pp. 1721–1735, 2019.
- [22] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested u-net architecture for medical image segmentation," 2018, *arXiv:1807.10165*. [Online]. Available: <https://arxiv.org/abs/1807.10165>
- [23] M. Zahangir Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*. [Online]. Available: <http://arxiv.org/abs/1802.06955>
- [24] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: Hints for thin deep nets," 2014, *arXiv:1412.6550*. [Online]. Available: <https://arxiv.org/abs/1412.6550>
- [25] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [27] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, Sep. 2010, pp. 249–256.
- [28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [29] S. Jegou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1175–1183.
- [30] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.
- [31] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 568–576, Feb. 2020.
- [32] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "DRINet for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2453–2462, Nov. 2018.
- [33] C. Iribarren, A. S. Go, I. Tolstykh, S. Sidney, S. C. Johnston, and D. B. Spring, "Breast vascular calcification and risk of coronary heart disease, stroke, and heart failure," *J. Women's Health*, vol. 13, no. 4, pp. 381–389, May 2004.
- [34] E. M. P. Fiuzza Ferreira, J. Szejnfeld, and S. Faintuch, "Correlation between intramammary arterial calcifications and CAD," *Academic Radiol.*, vol. 14, no. 2, pp. 144–150, Feb. 2007.
- [35] B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based segmentation of medical imaging data," 2017, *arXiv:1701.03056*. [Online]. Available: <http://arxiv.org/abs/1701.03056>
- [36] M. Zhai, J. Liu, W. Zhang, C. Liu, W. Li, and Y. Cao, "Multi-scale feature fusion single shot object detector based on DenseNet," in *Intelligent Robotics and Applications*, H. Yu, J. Liu, L. Liu, Z. Ju, Y. Liu, and D. Zhou, Eds. Cham, Switzerland: Springer, 2019, pp. 450–460.
- [37] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1933–1941.
- [38] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2999–3007.
- [39] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, "The digital database for screening mammography," in *Proceedings 4th Int. Workshop Digit. Mammography*, 2000, pp. 212–218.
- [40] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Sci. Data*, vol. 4, no. 1, pp. 170–177, Dec. 2017.
- [41] *Labelbox*. Accessed: Jan. 1, 2019. [Online]. Available: <https://labelbox.com>
- [42] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," *Biochimica et Biophys. Acta (BBA)-Protein Struct.*, vol. 405, no. 2, pp. 442–451, Oct. 1975.
- [43] D. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, pp. 2229–3981, Oct. 2011.
- [44] D. P. Chakraborty and L. H. Winter, "Free-response methodology: Alternate analysis and a new observer-performance experiment," *Radiology*, vol. 174, no. 3, pp. 873–881, Mar. 1990.
- [45] F. Chollet (2015). *et al. Keras*. [Online]. Available: <https://keras.io>
- [46] L. Bottou, *Stochastic Gradient Descent Tricks*. Berlin, Germany: Springer, 2012, pp. 421–436.
- [47] H. D. Menéndez and J. L. Llorente, "Mimicking anti-viruses with machine learning and entropy profiles," *Entropy*, vol. 21, no. 5, p. 513, 2019.
- [48] F. Angiulli, "Prototype-based domain description for one-class classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1131–1144, Jun. 2012.