IEEE Transactions on Medical Imaging Volume: 38 , Issue: 1 , Jan 2019

Tumor Detection in Automated Breast Ultrasound Using 3-D CNN and Prioritized Candidate Aggregation

Tsung-Chen Chiang, Yao-Sian Huang, Rong-Tai Chen, Chiun-Sheng Huang*, and Ruey-Feng Chang*, *Senior Member, IEEE*

***Abstract*—Automated whole breast ultrasound (ABUS) has been widely used as a screening modality for examination of breast abnormalities. Reviewing hundreds of slices produced by ABUS, however, is time-consuming. Therefore, in this study, a fast and effective computer-aided detection (CADe) system based on 3-D convolutional neural networks (CNN) and prioritized candidate aggregation is proposed to accelerate this reviewing. Firstly, an efficient sliding window method is used to extract volumes of interest (VOIs). Then, each VOI is estimated the tumor probability with a 3-D CNN, and VOIs with higher estimated probability are selected as tumor candidates. Since the candidates may overlap each other, a novel scheme is designed to aggregate the overlapped candidates. During the aggregation, candidates are prioritized based on estimated tumor probability to alleviate over-aggregation issue. The relationship between the sizes of VOI and target tumor is optimally exploited to effectively perform each stage of our detection algorithm. On evaluation with a test set of 171 tumors, our method achieved sensitivities of 95% (162/171), 90% (154/171), 85% (145/171), and 80% (137/171) with 14.03, 6.92, 4.91, and 3.62 FPs per patient (with 6 passes), respectively. In summary, our method is more general and much faster than preliminary works, and demonstrates promising results.**

***Index Terms*—Automated whole breast ultrasound, breast cancer, computer-aided detection, convolutional neural networks.**

I. Introduction

BREAST cancer is the second leading cause of death for women in the world [1]. Early detection and treatment are important in reducing mortality rates [2]. Ultrasonography is widely used in detection and diagnosis of breast tumors. Conventionally, 2-D handheld breast ultrasound (US) [3, 4] was used as an adjunct modality to the mammography [5, 6]. Nevertheless, the handheld US is time-consuming, operator dependent, and has poor reproducibility. To overcome these limitations, the automated whole breast ultrasound (ABUS) has been proposed to scan the whole breast automatically. The breast cancer detection rates are significantly increased using ABUS in conjunction with mammography for women with dense breast tissues [7, 8]. In contrast to handheld US, the ABUS is less operator dependent and had demonstrated greater reproducibility for follow-up studies [9]. Besides, the ABUS produces 2-D slices that can be reconstructed to 3-D volume for further review. A sample ABUS slice with a lesion is illustrated in Fig. 1. Reviewing hundreds of slices produced by the ABUS, however, requires a large amount of time even for expert physicians. In order to reduce the reviewing time, several computer-aided detection (CADe) systems had thus been proposed to assist in the reviewing and forming more accurate detection and diagnosis for ABUS images [7, 8, 10-13].

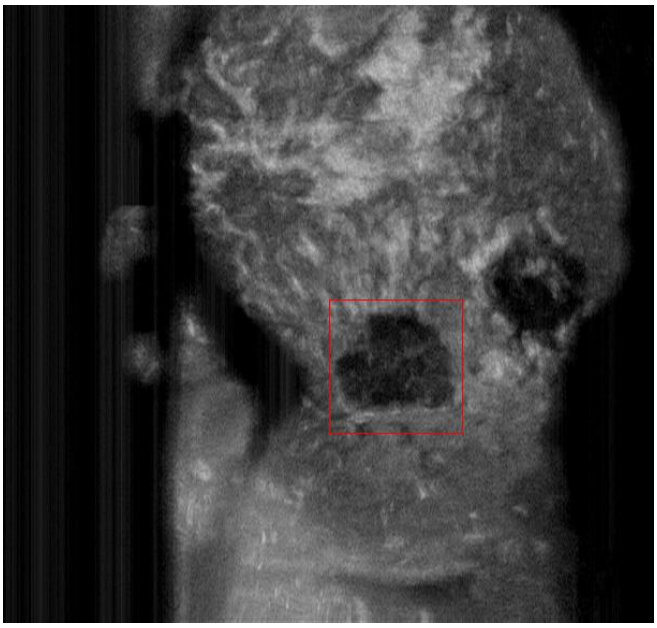


Fig. 1. Illustration of the ABUS image and the tumor detection problem. A ground truth tumor is indicated in the red box.

A fully automatic scheme for mass detection was developed by Ikedo *et al*. [8]. In this method, edges were detected using the Canny edge detector. Near-vertical edges and near-horizontal edges were discriminated, and the near-vertical edges were considered as potential tumor positions. Then, the watershed transform was performed for segmentation of the

This work was supported by the Ministry of Science and Technology (MOST 107-2634-F-002-013, MOST 107-2634-F-002-019) of the Republic of China for the financial support. *Asterisk indicates corresponding author.*

T.-C. Chiang, Y.-S. Huang, and R.-T. Chen are with the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan.

*C.-S. Huang is with the Department of Surgery, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan (e-mail: huangcs@ntu.edu.tw).

*R.-F. Chang is with the Department of Computer Science and Information Engineering, the Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, MOST Joint Research Center for AI Technology and All Vista Healthcare, Taipei, Taiwan (e-mail: rfchang@csie.ntu.edu.tw).

located positions and generated the regions of tumor candidates. Chang *et al*. [7] proposed a CADe system to detect breast lesions in multi-pass automated breast US. Firstly, the images were pre-processed. Then, the tumor candidates were segmented with the gray level slicing method. Finally, seven quantitative features were extracted for discrimination between tumors and non-tumors. Tan *et al*. [13] proposed a multi-stage CADe system including segmentations of the breast, the nipple, and the chest-wall, followed by voxel features extraction, and distinguishing between tumors and non-tumors using an ensemble of neural network classifiers. Moon *et al*. [12] proposed a method based on multi-scale blob detection, followed by a logistic regression (LR) classifier using blobness, internal echo, and morphology features to reduce the number of false positives (FPs). Another approach suggested by Lo *et al*. [11] pre-processed 2-D images, applied watershed transform to get homogeneous regions, and estimated the probabilities of candidates being tumors using the 2-D and 3-D texture, intensity, and morphology features with a LR classifier. Nevertheless, these approaches have the following drawbacks. First, the tumor candidate proposal schemes may be less generic and may over-fit the data set. Second, the selection of hand-crafted features requires specialized domain knowledge, which is inconvenient and may not be optimal. Third, these methods have relatively insufficient performance for clinical trials in terms of both detection rate and execution time.

To overcome these issues, an efficient algorithm for the ABUS tumor detection is proposed in this study. Instead of hypothesizing tumor candidates, the sliding window detector, which is another commonly used method for object detection [14, 15], is adopted to extract volumes of interest (VOIs) uniformly around the volume because of its generality and application independence. Thereafter, a 3-D convolutional neural networks (CNNs) is used for tumor probability estimation of each VOI. As opposed to hand-crafted features, the explicit definitions of feature designs can be avoided using the CNNs. The CNNs are able to learn a hierarchy of increasingly complicated features automatically and directly from a large amount of data. The learnt features are automatically optimized to fit the provided data set. Therefore, the focus of CNNs is on designing architectures. The CNNs have presented more promising results in several object recognition tasks including the handwriting digits recognition [16] and the ImageNet challenge [17]. In recent years, the CNNs were also gaining more popularity in analyzing medical images acquired by various modalities including detection tasks [18, 19]. Furthermore, to our best knowledge, the 3-D CNN has not been applied in ABUS tumor detection. Finally, a novel aggregation scheme is proposed for combination of overlapped VOIs with higher tumor probability.

This paper is organized as follows. In Section II, the used data acquisition and information of lesions are presented. Section III provides detailed description of the method and the evaluation metrics. The experimental results are presented and discussed in Section IV and Section V, respectively. Finally, Section VI presents the main conclusion.

TABLE I
THE DISTRIBUTION OF DIFFERENT LESION TYPES AND BI-RADS BREAST DENSITY TYPES IN EACH SET

		Training (25 patients, 29 tumors)	Validation (25 patients, 30 tumors)	Test (137 patients, 171 tumors)	Total
Tumor type	Fibrocystic Change	6	7	37	50
	Fibroadenoma	8	5	20	33
	Papilloma	2	0	5	7
	IDC	11	12	94	117
	DCIS	2	6	15	23
BI-RADS density type	Type A	2	0	10	12
	Type B	8	11	64	83
	Type C	12	13	47	72
	Type D	3	1	16	20

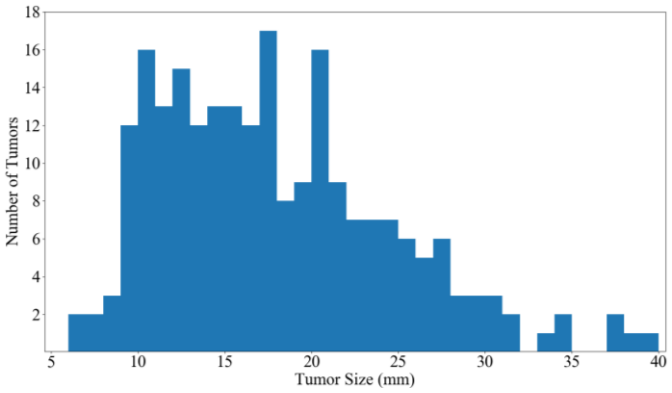


Fig. 2. Histogram of lesion size distribution of all 187 patients.

II. MATERIALS

ABUS images used in this study were acquired between January and September 2015 in the Breast Center of National Taiwan University Hospital from an ACUSON S2000 Automated Breast Volume Scanner (Siemens Medical Solutions, Mountain View, CA, USA) with a 14L5BV linear array transducer ranging from 5 to 15 MHz. The ABUS scanner produced 318 2-D images with 0.5 mm thickness. The pixel spacing along posterior-anterior and left-right axes are respectively 0.07 and 0.21 mm per pixel. These are standard ABUS views made for diagnostic assessment. To completely cover the breast, each patient was scanned in three passes for each breast. For each patient, all six passes were collected in our dataset. The informed consent of data usage in this retrospective study has been obtained from the institutional review board.

There are 230 pathology-proven lesions from 187 patients in our dataset, including 90 benign and 140 malignant lesions. Each ground truth tumor was labeled by the physician using a bounding box around the tumor. In benign lesions, 50 fibrocystic changes, 33 fibroadenomas, and 7 papillomas are included. The malignant lesions include 117 invasive ductal carcinomas (IDC) and 23 ductal carcinomas in situ (DCIS). Firstly, the data were randomly shuffled. Then, the first 25 patients were assigned to the training set, the next 25 patients to the validation set, and the rest 137 patients to the test set. In the literature of machine learning, training set is used to train the classifier and adjust the parameters (e.g., weights and biases of

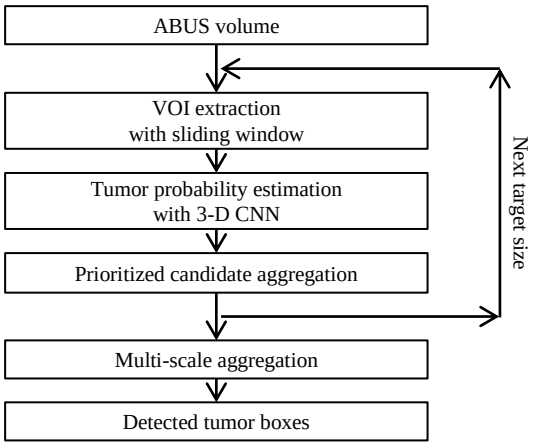


Fig. 3. Schematic flowchart of the proposed CADe system.

neural networks) within the model; validation set is used to tune the hyper-parameters, which refers to the exterior parameters (e.g., algorithm design and model selection) set prior to the commencement of the learning process; test set is used to evaluate the final solution and verify that the hyper-parameters do not over-fit the validation set. The distribution of lesion types and breast density types using Breast Imaging Reporting and Data System (BI-RADS) [20] in each set is listed in Table I. The histogram of lesion size is illustrated in Fig. 2. In addition to the 187 abnormal cases, 37 normal ABUS cases (each with 6 passes) without biopsy-proven tumors were also collected in test set.

III. Methods

In this study, a fast and effective CADe system based on 3-D CNN is proposed for breast tumor detection in 3-D ABUS. Our proposed detection algorithm takes as parameters a list of target tumor sizes (*L*s) and the degree of aggregation (*DoA*). The algorithm involves three main stages: the VOI extraction, tumor probability estimation with the 3-D CNN, and the candidate aggregation. At first, an efficient 3-D sliding window method is used to extract the VOIs. Then, the 3-D CNN is used to estimate the probability being tumor of each VOI, and the VOIs with tumor probability greater than a threshold are selected as tumor candidates. However, some of the candidates may overlap each other. Hence, a candidate aggregation method based on the hierarchical clustering [21, 22] is proposed to combine the overlapped candidates into a single tumor box, where each candidate is scheduled with different priority for alleviating the over-aggregation problem. Finally, to detect lesions of different sizes, the aforementioned steps are performed multiple times at different scales, and a simple scheme is adopted for multi-scale tumor VOI aggregation. Fig. 3 illustrates the schematic flowchart of the proposed CADe system. In this section, each step of the detection algorithm will be described in details.

A. VOI Extraction with Sliding Window

For VOI extraction, the first step of our CADe system employs the 3-D sliding window to scan the whole ABUS volume. When the sliding window moves with a stride, a VOI will be extracted. For a target size *L*, our CADe system uses the stride *L* to extract VOIs of size 2*L* for the following reason. Although the tumor can be entirely covered with higher probability using a small stride, the number of extracted VOIs will be very large. In fact, a tumor of size less than or equal to *L* is guaranteed to be completely covered by at least one VOI using a sliding window of size 2*L* and stride $\leq L$. Therefore, the settings will reduce the execution time while simultaneously produce VOIs covering the entire target tumors. Furthermore, since the 3-D CNN requires inputs of the same dimension, each VOI will be rescaled to the same size 32×32×32.

TABLE II
THE ARCHITECTURE OF PROPOSED 3-D CNN

Type	# kernels	Kernel size	Stride	# nodes	Input size	Dropout prob.
Conv.	32	5×5×5	1×1×1	-	32×32×32	-
Conv.	32	5×5×5	1×1×1	-	32×32×32×32	-
Max-pool.	-	2×2×2	2×2×2	-	32×32×32×32	-
Conv.	64	5×5×5	1×1×1	-	32×16×16×16	-
Conv.	64	5×5×5	1×1×1	-	64×16×16×16	-
Max-pool.	-	2×2×2	2×2×2	-	64×16×16×16	-
FC	-	-	-	128	32768	0.5
FC	-	-	-	64	128	0.5
FC	-	-	-	32	64	0.5
FC	-	-	-	2	32	-

Note. Conv., Max-pool., and FC stand for convolutional layers, max-pooling, and fully-connected, respectively.

TABLE III
THE HYPER-PARAMETERS OF PROPOSED CNN ARCHITECTURE

Stage	Hyper-parameter	Value
Initialization	Weights	Rand. uniform -0.05 ~ 0.05
	Bias	0.0
LeakyReLU	α	0.3
Adam Optimizer	Learning rate	0.001
	β_1	0.9
	β_2	0.999
	ε	1.0×10^{-8}
Training	Epochs	300
	Batch size	25

B. Tumor Probability Estimation with 3-D CNN

After the VOI extraction, each VOI will be estimated the probability being tumor by the CNN. One way to deal with 3-D data is to use 2-D CNN to predict each 2-D slice, and combine the results using recurrent neural network techniques such as long short-term memory (LSTM) [23]. Recurrent neural networks work by caching an internal state of the network that allows it to simulate temporal behavior, which mimics the way a human physician observes an ABUS volume by considering adjacent 2-D slices to find relations in sequences. However, the 3-D CNN has the ability to extract 3-D features, which more directly include information of relationship between adjacent voxels from arbitrary directions. Therefore, 3-D CNN is adopted in this study. As to the CNN architecture, several complicated designs such as AlexNet [17], VGGNet-16 [24],

and ResNet-34 [25] had been proposed and showed excellent results in classification of natural images. However, the 3-D version of those deep neural network models contain too many parameters, and the processing time will increase dramatically. Since medical data meet certain acquisition criteria [18], a CNN with less trainable parameters suffices. Therefore, a simplified 3-D CNN architecture was designed in this study. The architecture and hyper-parameters of the 3-D CNN are depicted in Table II and Table III, respectively.

As shown in Table II, the proposed 3-D CNN architecture consists of four convolutional layers with the number of kernels 32, 32, 64, and 64, respectively. The outputs of the second and fourth convolutional layers are down-sampled by the following 2×2×2 max-pooling layers. The objective of max pooling is to reduce computational cost and over-fitting by providing a more abstracted form of the input image. In second max-pooling layer, the outputs connect to a neural network consisting of three fully-connected (FC) layers for binary classification. Furthermore, for decreasing the effect of over-fitting, the Dropout [26], which removes the neuron from the network with probability p during training, is adopted for regularization in the FC layers. It behaves as a regularizer by preventing neurons from co-adapting to each other. In artificial neural networks, the activating function of a neuron defines the output of that neuron given an input or set of inputs. The rectified linear units (ReLU) [27], defined as $f(x) = \max(0, x)$, have been widely used as activating function for additional speed-ups, as opposed to conventional function such as sigmoid and hyperbolic tangent functions. However, the dying ReLU problem will occur when a ReLU neuron is pushed into a state in which the gradient becomes zero. The leaky rectified linear units (LeakyReLU) [28] is a variant of ReLU to cope with this problem by introducing a small slope when the neuron is not active. The LeakyReLU is defined as $f(x) = \max(0, x) + \alpha \cdot \min(0, x)$, where α is the leakage coefficient. Except for the output layer with sigmoid function for binary classification, all the other layers use LeakyReLU as the activating functions. Moreover, Adam [29] is used as the stochastic optimization solver for improving the speed in training CNN, and the cross-entropy is used as the loss function to optimize the weights and biases.

1) Training the 3-D CNN

In order to train the 3-D CNN, the VOIs of tumor and non-tumor class have to be provided. A VOI of any size extracted from anywhere in the ABUS can be assigned to non-tumor class as long as it satisfies the following two conditions. First, the center of the VOI is not included by any ground truth tumor box. Second, the VOI does not include the center of any tumor. In this study, the sizes of the extracted non-tumor VOIs are randomly selected in the range from 5 to 40 mm.

For the tumor class, however, it is unsuitable to simply use those VOIs that failed to be assigned to non-tumor as tumor VOIs, because many of these VOIs exhibit less essential tumor characteristics. For instance, these VOIs may crop the tumor overmuch, or only cover a small part of the tumor. Training the network with the most representative data conforming to the features for each class can decrease the convergence time and improves the generalization ability. Hence, instead of using the VOIs extracted by the sliding window, the ground truth tumor positions and sizes were used to accurately extract the tumor VOIs to train the 3-D CNN. The size of the tumor VOIs is two times the size of the tumor, which is consistent with the relative size between the target tumors and the sliding window. Only one 3-D CNN was trained using all tumor and non-tumor VOIs of all sizes.

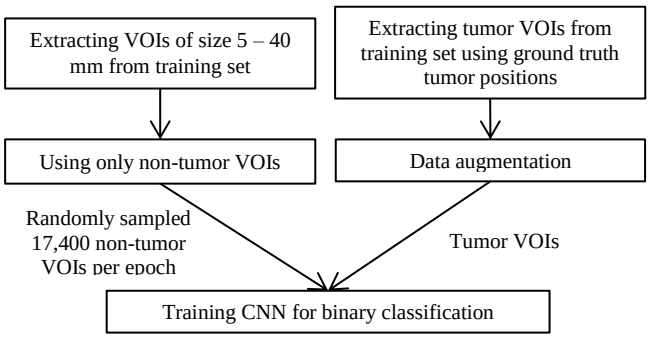


Fig. 4. Flowchart of training CNN.

TABLE IV
THE NUMBER OF TUMOR AND NON-TUMOR VOIS IN TRAINING AND VALIDATION SET AFTER DATA AUGMENTATION

	Training set	Validation set
Non-tumor	~400,000	~400,000
Tumor	17,400	18,000

In addition, since the number of tumor VOIs is much less than that of non-tumor VOIs, data augmentation is applied to the tumor VOIs. Data augmentation not only increases the number of training data but also improves the robustness and generalization of the CNN. Therefore, 100 times of shifting (± 20% relative to the VOI size along three orthogonal axes), scaling (± 20%), and flipping (along superior-inferior and left-right directions) are randomly applied to each tumor VOI. After data augmentation, the corresponding number of training and validation VOIs for each class is listed in Table IV. However, extreme unbalance still exists between each class. An unbalanced dataset will bias the model towards the more commonly emerged class. Hence, at the beginning of each training epoch, a subset of non-tumor data will be randomly sampled to match the number of tumor data during training. The flowchart of training the 3-D CNN is illustrated in Fig. 4.

C. Prioritized Candidate Aggregation

After tumor probability estimation using the 3-D CNN, the tumor candidates are selected from VOIs with probability higher than a threshold *TH*. However, a tumor will probably be covered by multiple overlapped candidates. The overlapped candidates should be aggregated into a single box. Therefore, a candidate aggregation algorithm based on the hierarchical clustering (HC) is proposed. In HC, a linkage criterion, which is a function of a dissimilarity metric, is used as the measure of dissimilarity between data sets. Two sets with dissimilarity less than a threshold will be combined into a cluster. The input of HC is the centers of tumor candidates and the parameters are listed in Table V. The HC will assign neighboring centers into

TABLE V
PARAMETERS OF HIERARCHICAL CLUSTERING

Parameter	Value
Dissimilarity metric	Euclidean distance $d(a, b) = \sqrt{\sum_i (a_i - b_i)^2}$
Linkage criterion	Single-linkage (nearest neighbor) $\min(d(a, b): a \in A, b \in B)$
Dissimilarity threshold	$\sqrt{3}L$

a cluster. The dissimilarity threshold used in HC is set to $\sqrt{3}L$, since the longest Euclidean distance between the centers of two VOIs of size $2L$ covering a tumor of size $\leq L$ is $\sqrt{3}L$. After clustering, the weighted average with the positions of candidates in the same cluster is computed and used as the position of the aggregated box where the weight assigned to each candidate is its estimated tumor probability. The size of the aggregated box remains $2L$.

1) Alleviation of Over-aggregation

When a lower threshold *TH* is applied for candidate selection, lots of candidates will emerge and densely distribute everywhere around the ABUS volume. As a result, the nearest neighbor criterion used in HC will group too many candidates into a cluster. Hence, a tumor box may be displaced from correct tumor position and cover the tumor incompletely. To address the issue, the cluster size (i.e. the number of candidates in a cluster) must be restricted and the candidates with higher tumor probabilities should be prioritized for aggregation. In particular, the highest threshold *TH'* > *TH* is first applied for candidate selection followed by the HC. Then, *TH'* is slightly decreased to select more candidates. However, a newly selected candidate will join its neighboring cluster only if the cluster has not reached the maximal size yet, where the maximal cluster size is referred to as degree of aggregation (*DoA*) in our algorithm. Because a tumor of size less than *L* will be entirely covered by at most eight VOIs of size $2L$, the optimized *DoA* should be no more than eight. Candidate selection and clustering are repeated for each *TH'* until *TH'* equals *TH*.

2) Multi-scale Aggregation

Because the above procedure only handles a single target tumor size *L*, our system allows the physician to input multiple target tumor sizes (multiple *L*s) and the same procedure will be performed multiple times on each target size for multi-scale tumor detection. As a consequence, the tumor boxes of multiple sizes may be produced and overlap each other. Since each tumor box has been obtained by aggregating multiple candidates, the maximal probability of the candidates within a cluster is used to represent the tumor probability of the aggregated tumor box. If a larger tumor box covers the centers of any smaller boxes, they will be aggregated using the weighted average with both the positions and sizes where the weights are the tumor probabilities.

D. Evaluation

For evaluating the performance of the proposed CADe system, the free-response receiver operating characteristics (FROC) curves [30] was adopted to evaluate the trade-offs between the sensitivity and number of FPs per patient (with 6 passes) at different tumor probability thresholds. A tumor box is determined to be a true positive if the distance between the center of the tumor box and the center of a true tumor is less than 10 mm; otherwise, the box is a false positive. For each FROC curve, area under the curve (AUC) was computed for FPs per patient under 12 (i.e., 2 FPs/pass); the corresponding sensitivities at 6 FPs/patient were marked; error bars were also computed using standard error of sensitivities at 1, 2, …, 12 FPs/patient.

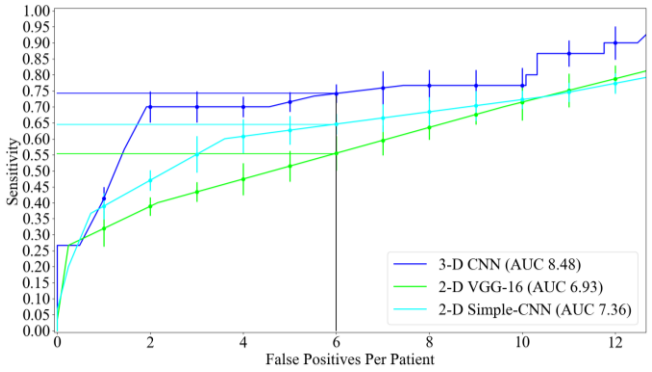


Fig. 5. FROC curves with error bars (using standard error) and the corresponding AUC for each CNN model evaluated on validation set.

TABLE VI
EXECUTION TIME (IN SECOND) OF OUR METHOD PER PATIENT (6 PASSES) WITH DIFFERENT CNNS AND TARGET TUMOR SIZES *L*S (IN MILLIMETER)

L (# VOIs)	Rescaling VOIs	Classification of boxes by CNN		
		2-D VGGNet-16	2-D Simple CNN	3-D CNN
7.5 (13200)	36	12	6	36
12.5 (2808)	24	~2	~1	6
17.5 (864)	18	~1	~1	~1

IV. EXPERIMENTS AND RESULTS

All experiments were accomplished using Theano framework [31] on a machine with an Intel Core i7-6700K 4.0 GHz processor and an NVIDIA GeForce GTX 1080 graphic card.

For comparison with deeper CNN architectures, a simple 2-D CNN (which had the similar architecture as our 3-D CNN described in Table II) and a 2-D VGGNet-16 were evaluated on validation set. 2-D CNNs were used instead of 3-D CNNs for time efficiency. The LSTM was used to adapt 2-D networks to our 3-D data. In subsequent paragraphs, the mention of 2-D CNNs implicitly indicates the usage of LSTM unless otherwise specified. To show the effectiveness of 3-D CNN, comparison between 3-D and 2-D CNNs was also performed on validation set. The FROC curves and the corresponding AUC of these CNN models are illustrated in Fig. 5. The execution time of each CNN model is also compared (Table VI). Since the 3-D CNN outperforms the other models in terms of FROC curve, it was selected as our classification model in all subsequent experiments on test set.

One of the parameters required by our detection algorithm is

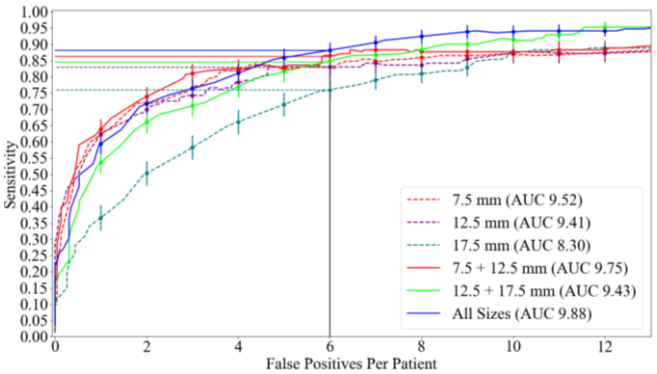


Fig. 6. FROC curves with error bars (using standard error) and the corresponding AUC using different target tumor sizes evaluated on test set. *DoA* of 4 was used for each curve.

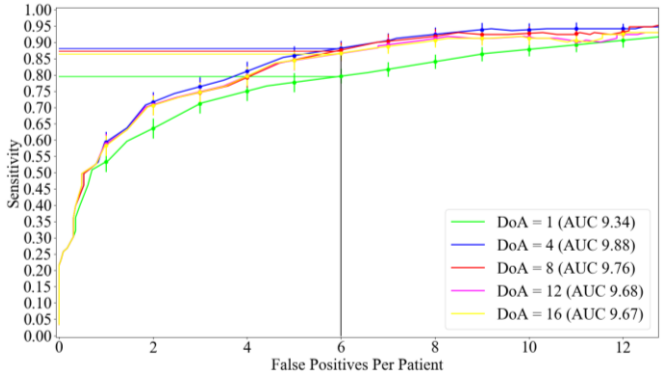


Fig. 7. FROC curves with error bars (using standard error) for comparison of different degrees of aggregation (*DoA*) evaluated on test set: *DoA*=1 (no aggregation), 2, 4, 8, 12, and 16. Three target sizes (7.5, 12.5, and 17.5 mm) were used for each curve.

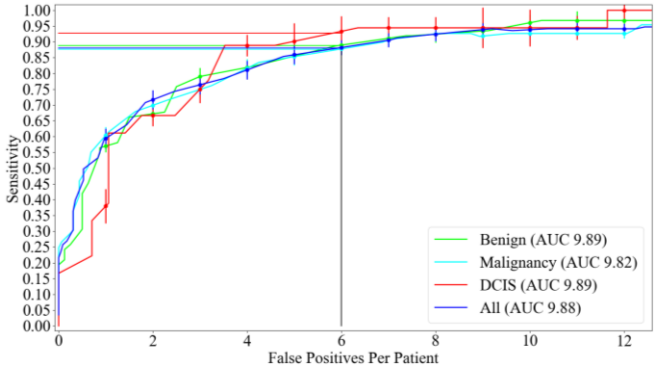


Fig. 8. FROC curves with error bars (using standard error) for benign, malignant, and DCIS lesions evaluated on test set. For each curve, *DoA* of 4 and three target sizes (7.5, 12.5, and 17.5 mm) were used.

a list of target tumor sizes (*L*s). Three target sizes were used in our experiments: 7.5, 12.5, and 17.5 mm. The sizes were selected to optimize the performance on validation set. Note that since the size in posterior-anterior direction of our ABUS volumes is smaller than 40 mm, the target size cannot be greater than 20 mm. The FROC curves and AUC using different combinations of the three target sizes on test set were computed to evaluate the effect of the parameter, as illustrated in Fig. 6. The number of extracted VOIs and the corresponding processing times for rescaling the extracted VOIs are also recorded in Table VI. Rescaling the VOIs and predicting the tumor probability by the 3-D CNN dominate the execution time of the algorithm and the rest of the steps together takes no more than 1 second on average. One of the essential part of the

TABLE VII
FPS PER PATIENT OF ABNORMAL AND NORMAL CASES AT DIFFERENT TUMOR PROBABILITY THRESHOLDS. EACH ROW COMPARES FPS/PATIENT USING THE SAME THRESHOLD. FOR ABNORMAL CASES, THE CORRESPONDING SENSITIVITY IS ALSO LISTED.

FPs/patient of abnormal data (sensitivity)	FPs/patient of normal data
3.62 (80%)	3.92
4.91 (85%)	4.78
6.92 (90%)	7.13
14.03 (95%)	13.52
24.38 (97%)	21.22

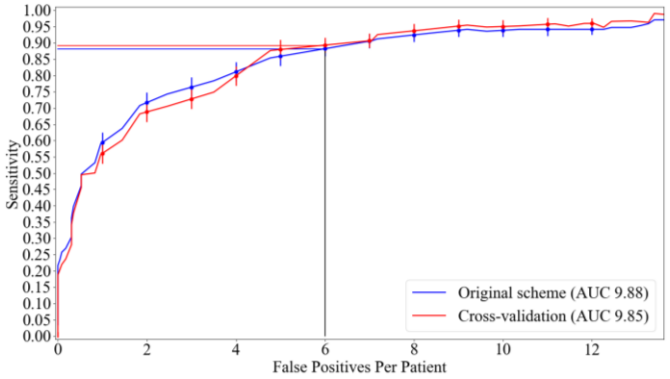


Fig. 9. FROC curves with error bars (using standard error) and the corresponding AUC using our original scheme and the 5-fold cross validation scheme evaluated on test set. For each curve, *DoA* of 4 and three target sizes (7.5, 12.5, and 17.5 mm) were used.

proposed algorithm is the aggregation process with a scheme to alleviate the over-aggregation phenomenon. The trade-offs between different restrictions on the cluster size (*DoA*) is illustrated in Fig. 7.

To examine the generalization of the algorithm, the number of FPs per patient of normal and abnormal cases at different tumor probability thresholds were also recorded (Table VII). The numbers of FPs for abnormal and normal cases do not show much difference at all thresholds. In addition, Fig. 8 compares FROC curves for benign, malignant, and DCIS lesions. Our CADe system achieves similar performance for each lesion type.

In addition to our original validation method, another experiment adopting a 5-fold cross-validation scheme was also performed. In each round of cross-validation, the dataset was partitioned into training (75 patients), validation (75 patients), and test set (37 patients). Different round of cross-validation contained completely different 37 patients for testing, and the rest of the 150 patients were randomly partitioned for training and validation. The average FROC curve on test set of the cross-validation scheme is illustrated in Fig. 9. The result is similar to our original scheme.

In order to illustrate the prediction result of the 3-D CNN, Fig. 10 visualizes the estimated probability of each extracted VOI in the blue channel using single target size L = 7.5 mm. In Fig. 11, the final detection results (after aggregation of multiple sizes) of two successful cases are presented. Fig. 12 shows two of the misdetection cases at sensitivity over 98%.

For comparison with previous methods, Table VIII lists detection results in terms of processing time and FPs per patient

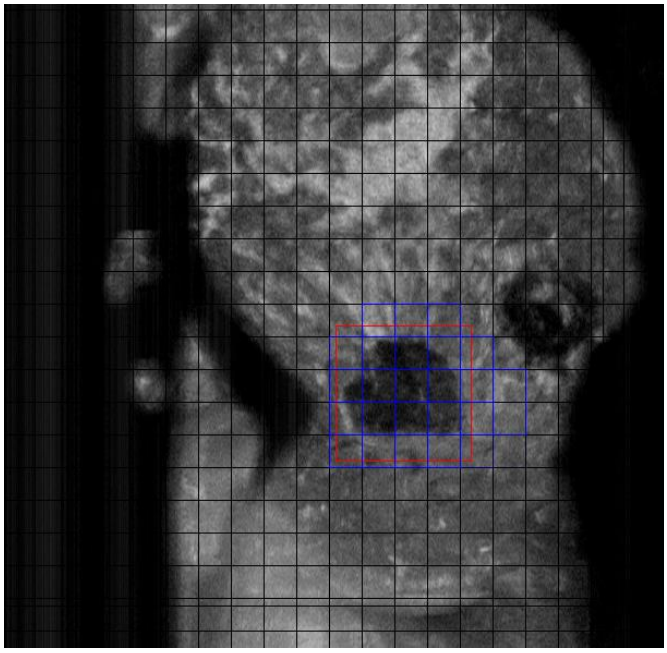
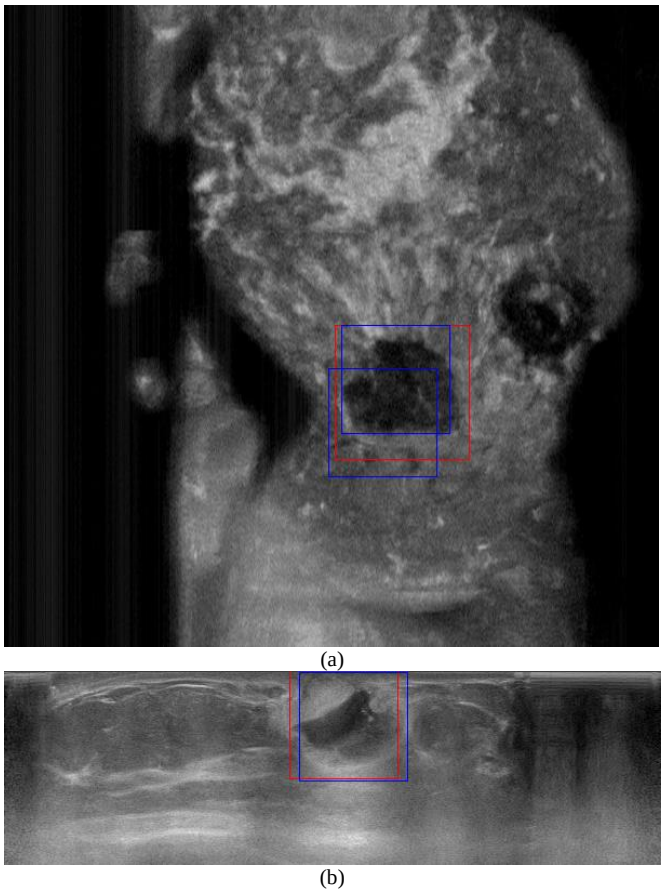


Fig. 10. Visualization of estimated lesion probability of each 15-mm VOI in blue channel. The case is a 31-mm DCIS with the ground truth indicated by the red box.

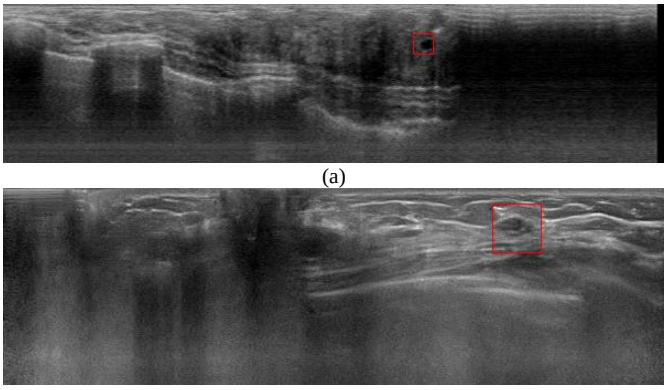


(a)

(b)

Fig. 11. True positive cases with boxes in red and blue respectively denoting the ground truth and results produced by our CADe system. (a) A 31-mm DCIS. (b) A 24-mm DCIS.

at different sensitivities reported in the recent literature on ABUS tumor detection. Note that the related works used



(a)

(b)

Fig. 12. The misdetection cases at 98% sensitivity. (a) A 6-mm fibrocystic change. (b) A 11-mm fibroadenomas.

TABLE VIII
SENSITIVITIES AND CORRESPONDING FPS/PATIENT OF OUR METHOD AND RELATED WORKS

Sensitivity (%)	**FPs/patient**		
	Our method	Lo *et al.* [11]	Moon *et al.* [12]
60.00	1.32	9.48	8.46
70.00	2.10	12.84	--
72.06	2.58	--	16.02
80.00	3.62	19.98	--
82.35	4.92	--	30.48
87.50	6.12	--	37.92
90.00	6.92	32.52	--
93.38	11.64	--	71.22
95.59	14.52	--	81.42
97.01	24.38	--	--
100.00	> 120.00	56.64	105.24
Number of tumors	171	133	136
Number of patients (abnormal + normal)	137 + 37	104 + 34	122 + 37
Execution time per patient	121 s	444 s	78 m

Note. Since the two compared methods reported results only in terms of FPs per pass instead of per patient (each with 6 passes in our dataset), these figures were multiplied by 6 to allow fair comparison.

different scans and the experiments were accomplished on different machines. Therefore, it is difficult to reliably compare the performance of different methods. Nevertheless, the results can still be used as an indication of differences among the methods.

V. DISCUSSION

A CADe system using an algorithm based on 3-D CNN and the prioritized candidate aggregation for ABUS tumor detection has been presented in this study. The detection algorithm was devised by exploiting the relationship between the size of target tumor and VOI to effectively perform each step: the stride (L) for sliding window, the dissimilarity threshold ($\sqrt{3}\,L$) for hierarchical clustering, and the criterion on the degree of aggregation ($DoA \leq 8$). On evaluation with a test set of 171 tumors, the CADe system demonstrated promising detection performance and time efficiency.

In contrast to previous works in ABUS tumor detection that hypothesize candidates using specifically designed image pre-processing methods [11-13], the sliding window detector is

more general and application independent. The drawbacks of using sliding window include the large number of extracted VOIs, the large amount of processing time, and being prone to produce more FPs. Nevertheless, the proposed algorithm effectively tackled the issues by using a large stride *L* for the sliding window to extract size-2*L* VOIs without the risk of missing target tumors.

After VOI extraction, each VOI has to be predicted the tumor probability using CNN. In comparison between the simple 2-D CNN and 2-D VGGNet-16 CNN (Fig. 5), our experiment indicates that a more complicated architecture does not contribute to the detection performance, and therefore a simpler architecture was adopted. Moreover, our proposed 3-D CNN outperforms both 2-D CNNs. As a result, a simple 3-D CNN is suggested in our study. Note that the number of trainable parameters of the 3-D CNN (5,105,026) is even much less than that of the 2-D VGGNet-16 (14,672,770), indicating that the 3-D CNN performs better by learning more essential features instead of fitting more parameters.

Due to the variable lesion sizes, our CADe system allows the physician to input a list of target sizes. The different sizes of sliding window aim at producing VOIs to bound tumors with different scales. Using a smaller sliding window can bound small tumors more compactly but may crop larger tumors and hence remove large-scale features such as shapes and blob-ness. On the other hand, a larger sliding window can cover more sizes, but small-scale features required to distinguish between speckle noises and small tumors are less recognizable. In the single target size comparison (dotted lines in Fig. 6), our system achieved better performance using a smaller size since a majority of our data consists of smaller tumors ($\leq$ 20 mm). Furthermore, even though a tumor may be cropped with small *L*, the 3-D CNN may still be able to recognize it, as illustrated in Fig. 10. On the contrary, using a larger *L* estimates most of the large tumors with higher probability but detects small tumors much less effectively. The result indicates that features relating to tumor margins are more crucial for identification of tumors. Moreover, the performance was further improved by aggregating results from multiple target sizes. The reason is that even for tumors of the same size, different tumors may be best recognized by the CNN at different VOI scales.

Another parameter of the proposed algorithm is the degree of aggregation *DoA*. With the *DoA* of 1, the algorithm equivalently performs no aggregation. As a result, all the candidates remain as the final tumor boxes. Without aggregation, however, a tumor may be bounded by several overlapped tumor boxes. Moreover, no FPs will be aggregated and therefore the number of FPs will increase dramatically at higher sensitivities. On the other hand, with higher *DoA*, the overlapped candidates will be combined into a single box. Hence, fewer FPs will be produced under the same sensitivity and a tumor can be bounded by fewer boxes. Besides, the aggregation process helps to achieve better localization. For instance, when a tumor of size *L* is entirely covered by eight candidate VOIs of size 2*L*, aggregating the candidates results in a single tumor box with the tumor located near its center. A potential problem of using higher *DoA* is over-aggregation, where incorrect candidates are combined with the correct ones. The jagged FROC curves in our experiments demonstrate the over-aggregating phenomenon, where decreasing the threshold *TH* to produce more candidates does not necessarily increase the sensitivity. The over-aggregation problem becomes more severe with *DoA* over eight because a size-*L* tumor can be completely covered by at most eight size-2*L* VOIs. Our study therefore suggests a compromise of *DoA* between 4 and 8.

In comparison with the related works, our method did not perform well at sensitivity higher than 98% and produced lots of FPs. Fig. 12 illustrates two misdetection cases at 98% (169/171) sensitivity. The failure reason of the fibrocystic change in Fig. 12 (a) probably comes from its small size and the plenty of surrounding shadow. The other failure case is a fibroadenomas, in Fig. 12 (b), which has obscured margins and shows less legible echoes, and the CNN therefore estimated it with a lower tumor probability. In spite of the limitation, our system outperforms the previous works distinctly at sensitivities under 98% and is considerably faster. Moreover, the previous works did not verify the generalization of their ad-hoc image processing methods to propose candidates (e.g., topographical watershed and blob-ness detection) and the hand-crafted features with a separate test set. Although their classifiers had been verified using cross-validation, the candidate proposal methods and the selected features might still over-fit the validation set. On the contrary, with a relatively small training and validation set for our method development, our reported result evaluated on a separate test set may be more reliable. Finally, a recent study reported that the mean ABUS interpretation time for radiologists of varied experience is less than 3 minutes per patient [32]. Thus, compared to the previous works, our method is much more feasible on clinical use considering execution time.

Previously, the CNNs had been adopted for object detection by using region proposals such as regions with CNN features (R-CNN) and its enhanced descendants [33-35]. In R-CNN, the selective search algorithm [36] was used for region proposals. However, the selective search was designed by optimally exploiting characteristics of natural 2-D images, it does not perform well in medical images with single color channel and high noise levels. In medical imaging, de Vos *et al*. [18] proposed an approach based on CNN for 3-D anatomical structure detection. In this method, each 2-D slice from three orthogonal directions is independently classified by a dedicated CNN to determine the presence of the target structure. Then, the detected 3-D bounding boxes are generated by intersection of all slices capturing the presence of the target. However, in our task, the tumors usually only occupy a relatively small area of a slice, and detecting the presence of tumor using the entire slice is less suited. Therefore, the sliding window detector is used in our study to analyze more localized VOIs, where an existing tumor will become more noticeable.

There are some drawbacks in our proposed CADe system. First, the localization of tumor boxes requires to be further strengthened. For instance, the detected boxes may be slightly displaced, crop a little tumor margin, or not bound the tumor compactly, as shown in Fig. 11. Second, our system does not

sketch the contour of the detected tumor. The tumor contour is among the most essential features for classification between benign and malignant masses. Our future works therefore include the investigation of segmentation methods such as conventional image analysis techniques (watershed, active contours, etc.) or, more recently, the fully convolutional networks (FCN) [37] for pixel-wise dense predictions in semantic segmentation. Finally, our detection algorithm takes target sizes *L*s and *DoA* as parameters. Although our experiments have suggested feasible values (i.e., L = 7.5, 12.5, and 17.5 mm; *DoA* = 4), the parameters may not require much tuning in clinical trials. Nevertheless, with a scheme for automatic parameter selection, the CADe system will be more robust and easy to use.

VI. Conclusion

A CADe system based on 3-D CNN for lesion detection of ABUS images is proposed in this study. An application-independent sliding window detector is adopted for VOI extraction. Then, a 3-D CNN is used for tumor probability estimation of each VOI, and VOIs of probability higher than a threshold are considered as tumor candidates. The overlapped candidates are combined with a novel aggregation scheme. Finally, the same process is executed multiple times with different target sizes for multi-scale lesion detection. The performance of our CADe system is evaluated with a database containing 171 lesions and 37 normal cases. The proposed CADe system achieves sensitivities of 95% (162/171), 90% (154/171), 85% (145/171), and 80% (137/171) with 14.03, 6.92, 4.91, and 3.62 FPs per patient (with 6 passes), respectively. The execution time is 21 seconds for each pass. The results demonstrate the feasibility of our method. The number of FPs at 100% sensitivity, however, should be further reduced. Methods for sketching tumor contours will be investigated as well.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA Cancer J Clin*, vol. 66, no. 1, pp. 7-30, Jan-Feb 2016.

[2] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA Cancer J Clin*, vol. 65, no. 2, pp. 87-108, Mar 2015.

[3] T. M. Kolb, J. Lichy, and J. H. Newhouse, "Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations," *Radiology*, vol. 225, no. 1, pp. 165-75, Oct 2002.

[4] E. A. Sickles, R. A. Filly, and P. W. Callen, "Benign breast lesions: ultrasound detection and diagnosis," *Radiology*, vol. 151, no. 2, pp. 467-70, May 1984.

[5] M. A. Roubidoux, M. A. Helvie, N. E. Lai, and C. Paramagul, "Bilateral breast cancer: early detection with mammography," *Radiology*, vol. 196, no. 2, pp. 427-31, Aug 1995.

[6] T. E. Wilson, M. A. Helvie, and D. A. August, "Breast cancer in the elderly patient: early detection with mammography," *Radiology*, vol. 190, no. 1, pp. 203-7, Jan 1994.

[7] R.-F. Chang, K.-C. Chang-Chien, H.-J. Chen, D.-R. Chen, E. Takada, and W. Kyung Moon, "Whole breast computer-aided screening using free-hand ultrasound," *International Congress Series*, vol. 1281, no. Complete, pp. 1075-1080, 2005.

[8] Y. Ikedo *et al.*, "Development of a fully automatic scheme for detection of masses in whole breast ultrasound images," *Med Phys*, vol. 34, no. 11, pp. 4378-88, Nov 2007.

[9] W. K. Moon *et al.*, "Comparative study of density analysis using automated whole breast ultrasound and MRI," *Medical Physics*, vol. 38, no. 1, pp. 382-389, 11/12 2011.

[10] R. F. Chang *et al.*, "Rapid image stitching and computer-aided detection for multipass automated breast ultrasound," *Med Phys*, vol. 37, no. 5, pp. 2063-73, May 2010.

[11] C. M. Lo *et al.*, "Multi-Dimensional Tumor Detection in Automated Whole Breast Ultrasound Using Topographic Watershed," *IEEE Transactions on Medical Imaging*, vol. 33, no. 7, pp. 1503-1511, 2014.

[12] W. K. Moon, Y. W. Shen, M. S. Bae, C. S. Huang, J. H. Chen, and R. F. Chang, "Computer-aided tumor detection based on multi-scale blob detection algorithm in automated breast ultrasound images," *IEEE Trans Med Imaging*, vol. 32, no. 7, pp. 1191-200, Jul 2013.

[13] T. Tan, B. Platel, R. Mus, L. Tabar, R. M. Mann, and N. Karssemeijer, "Computer-aided detection of cancer in automated 3-D breast ultrasound," *IEEE Trans Med Imaging*, vol. 32, no. 9, pp. 1698-706, Sep 2013.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 1, pp. 886-893: IEEE.

[15] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8: IEEE.

[16] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," presented at the Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 2012.

[18] B. de Vos, J. Wolterink, P. de Jong, T. Leiner, M. Viergever, and I. Isgum, "ConvNet-Based Localization of Anatomical Structures in 3D Medical Images," *IEEE Transactions on Medical Imaging*, 2017.

[19] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60-88, 2017.

[20] C. D'orsi, L. Bassett, W. Berg, S. Feig, V. Jackson, and D. Kopans, "Breast imaging reporting and data system: ACR BI-RADS-mammography," *American College of Radiology*, vol. 4, 2013.

[21] W. H. Day and H. Edelsbrunner, "Efficient algorithms for agglomerative hierarchical clustering methods," *Journal of classification*, vol. 1, no. 1, pp. 7-24, 1984.

[22] R. Sibson, "SLINK: an optimally efficient algorithm for the single-link cluster method," *The computer journal*, vol. 16, no. 1, pp. 30-34, 1973.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[24] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR*, vol. abs/1409.1556, 2014.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *CoRR*, vol. abs/1512.03385, 2015.

[26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929-1958, 2014.

[27] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807-814.

[28] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, 2013, vol. 30, no. 1.

[29] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[30] D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Medical physics*, vol. 16, no. 4, pp. 561-568, 1989.

[31] T. D. Team, "Theano: A Python framework for fast computation of mathematical expressions," *arXiv e-prints*, vol. abs/1605.02688, 2016.

[32] A. I. Huppe *et al.*, "Automated Breast Ultrasound Interpretation Times: A Reader Performance Study," *Academic radiology*, 2018.

[33] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.

[34] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91-99.

[35] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings*

of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587.

[36] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International journal of computer vision*, vol. 104, no. 2, pp. 154-171, 2013.

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