International Journal of Medical Informatics 117 (2018) 44–54

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification

Mugahed A. Al-antari[a,1], Mohammed A. Al-masni[a,1], Mun-Taek Choi[b], Seung-Moo Han[a], Tae-Seong Kim[a,*]

[a] Department of Biomedical Engineering, College of Electronics and Information, Kyung Hee University, Yongin, 17104, Republic of Korea
[b] School of Mechanical Engineering, Sungkyunkwan University, Republic of Korea

ARTICLE INFO

Keywords:
Computer-aided diagnosis (CAD)
Mass detection
You-only-look-once (YOLO)
Mass segmentation
Full resolution convolutional network (FrCN)
Deep learning

ABSTRACT

A computer-aided diagnosis (CAD) system requires detection, segmentation, and classification in one framework to assist radiologists efficiently in an accurate diagnosis. In this paper, a completely integrated CAD system is proposed to screen digital X-ray mammograms involving detection, segmentation, and classification of breast masses via deep learning methodologies.

In this work, to detect breast mass from entire mammograms, You-Only-Look-Once (YOLO), a regional deep learning approach, is used. To segment the mass, *full resolution convolutional network (FrCN)*, a new deep network model, is proposed and utilized. Finally, a deep *convolutional neural network (CNN)* is used to recognize the mass and classify it as either benign or malignant. To evaluate the proposed integrated CAD system in terms of the accuracies of detection, segmentation, and classification, the publicly available and annotated INbreast database was utilized. The evaluation results of the proposed CAD system via four-fold cross-validation tests show that a mass detection accuracy of 98.96%, Matthews correlation coefficient (MCC) of 97.62%, and F1-score of 99.24% are achieved with the INbreast dataset. Moreover, the mass segmentation results via FrCN produced an overall accuracy of 92.97%, MCC of 85.93%, and Dice (F1-score) of 92.69% and Jaccard similarity coefficient metrics of 86.37%, respectively. The detected and segmented masses were classified via CNN and achieved an overall accuracy of 95.64%, AUC of 94.78%, MCC of 89.91%, and F1-score of 96.84%, respectively. Our results demonstrate that the proposed CAD system, through all stages of detection, segmentation, and classification, outperforms the latest conventional deep learning methodologies. Our proposed CAD system could be used to assist radiologists in all stages of detection, segmentation, and classification of breast masses.

1. Introduction

Breast cancer is considered to be one of the most common types of cancer affecting women worldwide. Statistical results published in 2017 categorized breast cancer among the highest levels of all other cancers, accounting for 30% of estimated new cases and 14% of deaths [1]. In 2008, the World Health Organization (WHO) reported that 13.7% of deaths among women worldwide was due to breast cancer [2]. Early detection of breast cancer is a critical requirement for reducing the mortality rate among women [2–5]. At present, digital X-ray mammography is the most reliable screening device for suspicious breast masses and microcalcifications in the early stages [3,4,6,7]. Indeed, women over 40 years old are encouraged by the National Cancer Institute (NCI) to undergo breast screening one or two times per year using both views of mammograms: mediolateral oblique (MLO) and cranio-caudal (CC) [8]. In the diagnosis of breast abnormalities, clinical experts classify suspicious masses as benign or malignant. This task presents a daily challenge for radiologists due to the huge number of mammograms as well as the time and effort to examine each view of a mammogram [4,9,10]. Thus, a tradeoff between sensitivity and specificity has been realized during the diagnosis process. Through the use of a second reading, either by other experts or by a computer-aided diagnosis (CAD) system, the overall accuracy and specificity of mass detection, segmentation, and classification could be improved [3,11] and false positive and negative cases reduced. A reliable and robust CAD system could be of significant assistance in clinical practices [12,13].

* Corresponding author at: Department of Biomedical Engineering, College of Electronics and Information, Kyung Hee University, 1732, Deogyeong-daero, Giheung-gu, Yongin-si, Gyeonggi-do, 17104, Republic of Korea.
E-mail addresses: en.mualshz@khu.ac.kr (M.A. Al-antari), m.almasani@khu.ac.kr (M.A. Al-masni), mtchoi@skku.edu (M.-T. Choi), smhan@khu.ac.kr (S.-M. Han), tskim@khu.ac.kr (T.-S. Kim).
[1] Authors contributed equally to this work

https://doi.org/10.1016/j.ijmedinf.2018.06.003
Received 5 April 2018; Received in revised form 22 May 2018; Accepted 6 June 2018

There have been active developments for such a breast CAD system in each specific area of detection, segmentation, and classification. However, there are few studies involving a completely integrated system.

Mass detection from breast images is considered an important preprocessing stage to detect potential regions (i.e., masses) for further analysis by a CAD system. In fact, the variation of the masses within the surrounding tissues in terms of texture, shape, size, as well as the location in mammograms, makes the detection task challenging [8,13,14]. The majority of conventional CAD systems rely on manually detected masses and their hand extracted features to recognize the suspicious masses as benign or malignant via conventional machine learning techniques [11,15–18]. So far, this practice has resulted in a rather significant number of false positives [19–23]. Recently, novel detection approaches based on deep learning were introduced into a CAD system to overcome the challenging tasks of mass detection from mammograms [10,19,24].

Breast mass segmentation also plays a crucial role in accurately extracting discriminative shape features of specific mass regions, while excluding surrounding tissues [19,25]. In fact, improving overall accuracy in addition to reducing false positive and negative rates by mass segmentation is a big challenge due to the strong association between the presence of masses and their irregularities in shape, size, and location with low contrast and ambiguous boundaries [6,26–28]. Many studies involving mass segmentation have utilized region growing, active contour, and Chan-Vese methods [6,26,27]. Unfortunately, these methods still lack performance in handling mass segmentation automatically, because the simple hand-crafted or semi-automatic features based on prior knowledge cannot deal with complex shape variations, as well as the different density distribution of the masses and their surrounding tissues [6]. Recently, a few studies based on deep learning models have offered a good alternative to other conventional segmentation methods, by automatically extracting deep high-level hierarchy features for mass segmentation directly from input raw data to avoid the problems of hand-crafting features [4,19,29,30].

The majority of CAD systems have been developed in the area of mass classification to distinguish breast cancer as either benign or malignant utilizing conventional machine learning classifiers [11,15,16,17,18,31]. To build such systems, a set of hand-crafted or semi-automatic features describing the characteristics of masses are required. These features must have a good discriminative power to distinguish between either benign or malignant mass abnormalities. In fact, the conventional CAD systems based on hand-crafted features suffer due to the high degree of similarity between mass vs. non-mass and benign vs. malignant breast tissues [18,31,32]. Alternatively, some new strategies based on deep learning have recently been proposed to handle the mass classification task [3,31]. These strategies can learn and extract deep high-level features from raw input data directly and achieve much better classification performance in comparison to the traditional approaches [3,4,10,19,33]. In addition, a hybrid CAD system based on a combination of hand-crafted and deep high-level features has presented good results in the mass classification of breast masses [29].

The main contribution of this study is a fully integrated CAD system including three deep learning stages (i.e., detection, segmentation, and classification). Also the newly proposed segmentation method of FrCN is presented for the masses of breast cancer. The main advantage of FrCN is to preserve the high resolution of feature maps. Especially for the edges of the objects, where FrCN learns the full resolution features of each pixel of the original input data to achieve more accurate pixel-to-pixel segmentation. This is achieved by removing the max-pooling and subsampling layers in the networks and enabling the convolutional layers to extract and learn the full resolution spatial features of the input image.

In this paper, a completely integrated CAD system is proposed based on deep learning to automatically detect, segment, and classify breast masses in a single framework. The rest of the paper is organized as follows. First, an automatic deep learning You-Only-Look-Once (YOLO)-based mass detection model is presented. Second, a newly proposed deep learning mass segmentation method, a full resolution convolutional network (FrCN), is proposed and compared against other existing state-of-the-art deep models. Finally, a deep learning convolutional neural network (CNN) classifier is presented to distinguish between benign or malignant detected and segmented masses. We validate the proposed integrated CAD system and compare it to the latest methodologies by utilizing the public INbreast database [34].

2. Literature review

Breast cancer diagnosis via a CAD system can be improved by using the deep high-level features of deep learning, which can represent the characteristics of the masses better [4,10,19]. Mass detection is an important stage in the CAD systems for breast cancer diagnosis [4,35]. It is a challenging problem and has not been fully resolved [19,35]. In general, manual mass detection was utilized in the CAD systems which used deep CNN to classify the masses as either benign or malignant [33,36,37]. In other CAD systems, these manually detected masses are directly fed into CNN to generate the integrated high-level deep features [10,19,37]. In some other CAD systems, the high-level deep features were extracted from the multiple layers of CNN, and then concatenated and fed into the classifier to distinguish between benign or malignant tissues [33]. Most of these CAD systems achieved better classification performance against the traditional machine learning techniques which depend on the hand-crafted features [23,24]. However, the automatic mass detection still remains as a challenge. The need to automatically detect breast abnormalities was addressed in several studies [10,19,24,38]. At present, few deep learning studies present automatic mass detection methods in CAD systems [14,19,24]. The preliminary mass detection results are presented utilizing deep learning YOLO technique using the digital database of mammography (DDSM) [9]. The detection performance via YOLO was better in comparison to other recently published detection methods [14,19]. A CAD system utilizing a deep belief network (DBN) was presented to analyze suspicious regions in mammograms [4]. In this work, for mass detection, adaptive thresholding and morphological operations were utilized achieving an overall detection accuracy of 86% [3,4]. In [35], a new deep model called region-based CNN (R-CNN) was proposed to automatically detect the masses of breast cancer [35]. The entire mammogram was divided into multiple patches to detect the masses locally. Then, R-CNN was trained to classify the detected regions as benign or malignant. In [14], another automatic method using a cascade of deep learning models was proposed for mass detection. This method involved four sequential steps to detect masses in breast abnormalities. First, a multi-scale deep belief network (mDBN) and Gaussian mixture classifier (GMC) was utilized to extract suspicious regions. Second, two-level cascade of R-CNN was used to reduce the false positive rate in these detected regions. All remaining regions were then fed again into a two-level cascade of a conditional random forest (CRF) classifier to enhance the process of false positive reduction. All potential regions that survived in the previous stages were combined using a connected component analysis (CCA) as a post-processing technique [14]. Finally, the refinement algorithm was utilized to improve the precision of mass detection [39]. This refinement algorithm was also implemented with deep R-CNN through two sequential steps [19]. First, Bayesian optimization was used to detect suspicious regions. Then, a deep structure of R-CNN based classifier was utilized to improve the scale and localization of the detected regions. Despite improving the results for automatic mass detection, challenges remained with the high complexities of memory, practical implementation, and long runtime.

Several conventional studies have segmented masses from X-ray mammography images. Growing regions based on gradient filters and simple edge detection have been widely used for mass segmentation [26,40,41]. Other studies have improved the results of mass

segmentation for CAD systems by utilizing active contour and Markov random field (MRF) methods [42–44]. However, all of these methods have limitations because they depend on the prior knowledge of the mass contour [6,45]. Recently, a few studies based on deep learning with CNN have been presented and achieved better segmentation performance for medical and semantic image segmentation [46–50]. These segmentation models are mostly based on deep learning CNN which is built by adapting and converting the functionality of a well-known VGG-16 network from a classification to a segmentation task. However, due to the multiple of max-pooling and subsampling layers, these models suffer from the loss of spatial resolution of the feature maps. In general, the max-pooling and subsampling layers in each block of VGG-16 reduce the size of extracted feature maps, eliminate the redundancy of features, and minimize the computation cost [51]. Also due to max-pooling and subsampling layers, the spatial resolution of feature maps will exponentially decrease. One of the mostly used CNN models for segmentation is the fully convolutional network (FCN) [46]. FCN model consists of two main stages called the encoder and decoder networks [46]. The encoder network involves convolutional, max-pooling, and sub-sampling layers before the fully connected (FC) layers of VGG-16. To prevent the loss of the spatial resolution caused by the multiple max-pooling and subsampling layers, the decoder network is built by replacing the FC layers with the deconvolutional and up-sampling layers. Although the up-sampling layers could be used to recover the spatial resolution, this leads low segmentation accuracy, especially for tiny objects [46]. In [52], the deep FCN model was utilized to segment skin lesions from dermoscopy images [52]. To increase the feature map resolution, Jaccard distance was utilized as a loss function instead of cross-entropy in the training process. However, the problem of resolution reduction in the feature maps has not been resolved. Inspired by the structure of FCN, a new segmentation model called U-Net was proposed to segment neural brain images obtained from the electron microscopy (EM) [47]. In segmentation via U-Net, the feature maps from each encoder were combined with the corresponding one in the decoder network. Then, up-sampling and deconvolutional operations were performed to overcome the resolution loss of feature maps due to the multiple max-pooling and subsampling layers [47]. In [48], another deep segmentation model for pixel-to-pixel semantic segmentation called SegNet was proposed. Similar to FCN, SegNet model consisted of two main stages which were called the encoder and decoder networks. Each of encoder and corresponding decoder networks involved thirteen convolutional layers but in a reverse style of structure [48]. Finally, a Softmax classifier was used to produce the final segmentation maps with the same resolution of the original input image. Despite the encouraging results of all these segmentation techniques, they have not yet been applied to the mass segmentation of breast images. To date, only a few attempts based on deep learning have been presented for mass segmentation from mammograms. A deep learning model based on a structured support vector machine (SSVM) was proposed where the manual masses are extracted depending on the mass prior contour of the ground truth [30]. Then, DBN two times with patch sizes of 3×3 and 5×5 was utilized to detect potential candidates from the masses. Meanwhile, the deep model of DBN is combined with a Gaussian mixture classifier (GMC) to perform the pixel-to-pixel segmentation task. Improved mass segmentation results were achieved utilizing this segmentation method resulting in Dice indices of 87% and 88% on the DDSM-BCRP [36] and INbreast [34] datasets, respectively. In [45], CNN was utilized for breast cancer mass segmentation by comparing two different models: deep model based on SSVM and the other deep model based on conditional random field (CRF) classifier [45]. After that, Chan-Vese active contour model was utilized as refinement to improve the precision of mass segmentation results [53].

Recently, a few integrated CAD systems based on deep learning have been developed to include the detection, segmentation, and classification of breast masses in three consecutive stages [19,24]. In [29], a hybrid CAD system was proposed based on a combination of deep and hand-crafted features. CRF was first trained to generate the likelihood image where its local optima are used as seed points to generate potential masses [29]. Around the location of the seed point, the potential masses were extracted and then CNN was utilized to generate the deep high-level features. For segmentation, the region growing and active contour methods were used to segment the potential masses. Then, the hand-crafted features were manually generated from these segmented masses. The hand-crafted features and deep features were combined together to build the hybrid CAD system [29]. This hybrid CAD system achieved an AUC of 94.10%, while the CAD system which only depended on deep features achieved 92.90%. In [19] and [24], a comprehensive CAD system for breast cancer analysis was proposed. For mass detection, a complex cascade structure of deep learning was utilized involving mDBN with GMC, two stages of R-CNN, two stages of CRF, and a refinement algorithm based on R-CNN [14,19]. For mass segmentation, another complex cascade of deep learning techniques was used involving two stages of DBN with CRF and a refinement method by Chan-Vese active contour [19,21,24,45]. For classification, a simplified version of CNN was pre-trained to classify the masses as either benign or malignant. Despite of successes of these CAD systems for breast cancer diagnosis, the remaining challenges still exist including high complexities of memory, practical implementation, and long runtime challenges.

3. Materials and methods

In this study, we present an integrated CAD system for breast cancer which includes detection, segmentation, and classification in a single framework. First, an automatic mass detection is presented based on deep learning YOLO. Then, a mass segmentation methodology based on a novel deep learning FrCN is proposed. Finally, we also propose an automatic mass classification based on CNN. A schematic diagram of the proposed CAD system is illustrated in Fig. 1.

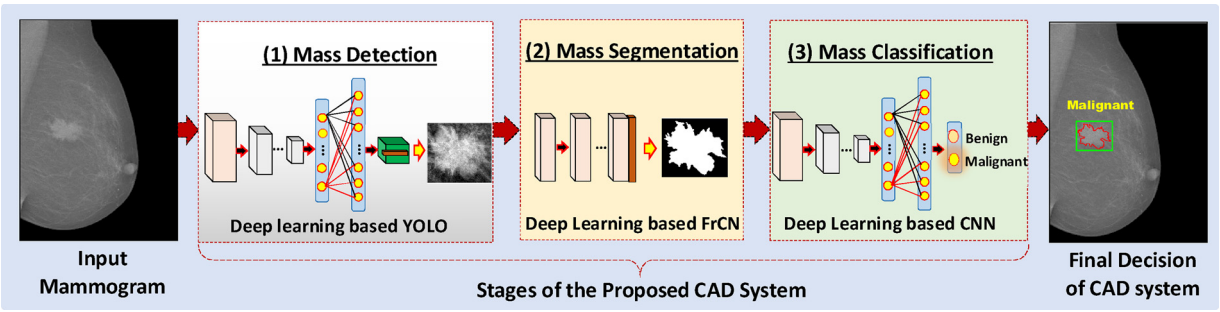

Fig. 1. Schematic diagram of the proposed computer-aided diagnosis (CAD) system based on deep learning to detect, segment, and classify breast cancer masses from input digital X-ray mammograms.

3.1. Dataset

In this study, the INbreast database [34] of X-ray mammography is utilized to test and evaluate our proposed CAD system. INbreast is the largest publicly available dataset with ground-truth annotations of breast cancer abnormalities (i.e., benign and malignant) [34]. It has 410 mammograms (i.e., normal, benign, and malignant) including views of both MLO and CC from 115 patients [34]. To evaluate our CAD system, we include all cases having masses in both views of the mammograms in a total of 107 cases [34]. Some of these cases have more than one mass, thereby, a total of 112 masses were collected according to the Breast Imaging Reporting and Data System (BI-RADS). BI-RAD is standard criteria developed by the American College of Radiology (ACR) to assign suspicious lesions into one of six categories [34]. Benign cases are assigned to the categories 2 and 3, while malignant cases are in categories 4, 5, and 6 [19,24]. In this study, 36 masses with BI-RAD $\in$ {2, 3} are categorized as benign, while 76 masses with BI-RAD $\in$ {4, 5, 6} are categorized as malignant.

3.1.1. Data augmentation and transfer learning

To train deep learning models, a large amount of annotated dataset is required. The small size of medical image datasets currently available presents a challenge for this training task [54]. Recently, two remedies are proposed to handle this challenge: *data augmentation* and *transfer learning*. Data augmentation is a well-accepted process that has been recently used to increase the size of the dataset, speed up the convergence, and avoid overfitting problems [10,19,24,33,54,55]. In this study, we have augmented the original mammograms eight times by rotating them with the angles of $\Delta\theta = 45^{\circ}$ (i.e., 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°) [10,19,24,29,33]. Thus, a total of 896 mammograms (i.e., the 112 original plus augmented mammograms) were collected to train and test all the proposed deep models: detection, segmentation, and classification. This means that in the total of 288 benign and 608 malignant cases were used in this study. For model initialization, there are two ways of initializing the parameters of deep models: random initialization and transfer learning [24,56]. In the former method, all parameters (i.e., weights) of deep models are randomly initialized with a zero-mean or unbiased Gaussian distribution with a standard deviation of 0.01 [57]. Meanwhile, some network biases of the convolutional and FC layers are initialized with number ones and others with zeros, where the initialization with ones accelerates the learning process by providing positive inputs to all activation functions (i.e., ReLUs) [57]. In the latter method, transfer learning is utilized to pre-train all deep models first with a large annotated computer vision dataset (i.e., ImageNet [58]) and then the models are re-trained or fine-tuned using the augmented annotated dataset (i.e., mammograms) [9,10,24,57]. In this work, we used the transfer learning method to initialize the parameters of all deep learning models. In fact, transfer learning has been utilized for breast cancer medical image analysis through various CAD systems [9,10,24,33].

3.2. YOLO for mass detection

Mass detection is the first critical task for the CAD system to detect the potential masses of breast tissues. It is a difficult task due to the large variation in the shape, size, and low SNR of masses from their surrounding tissues [9,19,29]. In this study, we have adopted a deep learning model called YOLO [9] to detect the mass regions (i.e., mass ROIs) from entire mammograms. Some preliminary work using YOLO has proven that this technique is effective in detection tasks [9,58]. YOLO is a regional ROI-based CNN technique established to directly detect suspicious regions of masses from the entire mammogram [9]. A good performance was achieved with YOLO for the detection of breast masses with a public X-ray mammography DDSM dataset [9]. YOLO can accurately detect and generate potential bounding boxes around breast masses [9]. Since the INbreast dataset includes accurate ground truth [19,24,34], YOLO can be a good choice for breast mass detection for the following reasons. First, YOLO has a robust ability to detect the masses directly from entire mammograms [9]. Second, detected bounding boxes via YOLO accurately align the masses, thereby, a low rate of false positives is achieved compared with other studies [14,19,24]. Third, it can even detect challenging cases where the masses exist either over pectoral muscles or inside dense regions. Fourth, the running time of the testing and required memory are extremely low compared to other more complex deep learning models [14,19,24].

3.3. FrCN for mass segmentation

Once the masses are detected from the previous detection stage of the CAD system, the detected masses are fed directly into our newly proposed segmentation stage. First, a contrast-limited adaptive histogram equalization (CLAHE) method is utilized as a preprocessing step for all detected masses. CLAHE is an image contrast enhancement algorithm which divides the entire image into multi-regions and then applies histogram equalization locally over each region [19,21,24,45]. This method has been successfully applied to improve image contrast and increase the contrast between the masses and their surrounding tissues [19,21,24,45,59].

Previous studies have established that FCN, SegNet, and U-Net for pixel-to-pixel segmentation which provided better segmentation results comparing to other conventional methods [46–48]. However, these segmentation models used multiple max-pooling and subsampling layers in their encoder networks resulting in the loss in the spatial resolution of the feature maps. To recover the resolution of the feature maps in the decoder network of these models, the up-sampling and deconvolution layers were utilized, but these processes increased the number of training parameters. These deep learning segmentation models suffer from the diminished spatial resolution, loss of details, and increase in computation cost.

In this study, we propose a new FrCN deep learning model for pixel-to-pixel mass segmentation. FrCN consists of two main consecutive encoder and decoder networks. The encoder network involves thirteen convolutional layers. However, unlike the previous deep models, the max-pooling and sub-sampling layers are removed from the encoder network to preserve the full spatial resolution of the original input as well as the details of the objects. This is a key modification to avoid any information loss during feature map generation for accurate pixel-to-pixel mass segmentation. Therefore, the high-level deep feature maps in each block of the encoder network are generated utilizing only the convolutional process, preserving the full resolution of the input images. By this modification, FrCN is able to maintain the details and edges especially for the tiny objects. Meanwhile, the decoder network of FrCN is built by replacing all three FC layers of VGG-16 with three full convolutional layers. Because the convolutional layers on the full resolution of the input images without sub-sampling in the encoder network are utilized, up-sampling and deconvolutional layers in the decoder network are not used. The final output of deep feature maps is directly fed into a Softmax classifier to obtain the probability that each pixel is a mass or non-mass. Finally, a non-linear activation function of ReLU is utilized after each block in the encoder and decoder stage as applied in previous work [48,49,51,57]. Indeed, the architecture of FrCN is inspired from recent applications in computer vision for semantic segmentation utilizing deep learning such as FCN [46], SegNet [48], DeconvNet [49], DeepLap [50], and U-Net [47]. Fig. 2 shows the architecture of the proposed deep learning FrCN segmentation model for pixel-to-pixel mass segmentation.

To establish our deep learning FrCN model, we utilize a stage-wise training process, where we gradually add decoder into the decoder network to develop the network until the performance is stable [46]. This deep model can successfully handle the segmentation task with full resolution and competitive computational time. To evaluate the overall segmentation performance of the proposed FrCN, a direct comparison

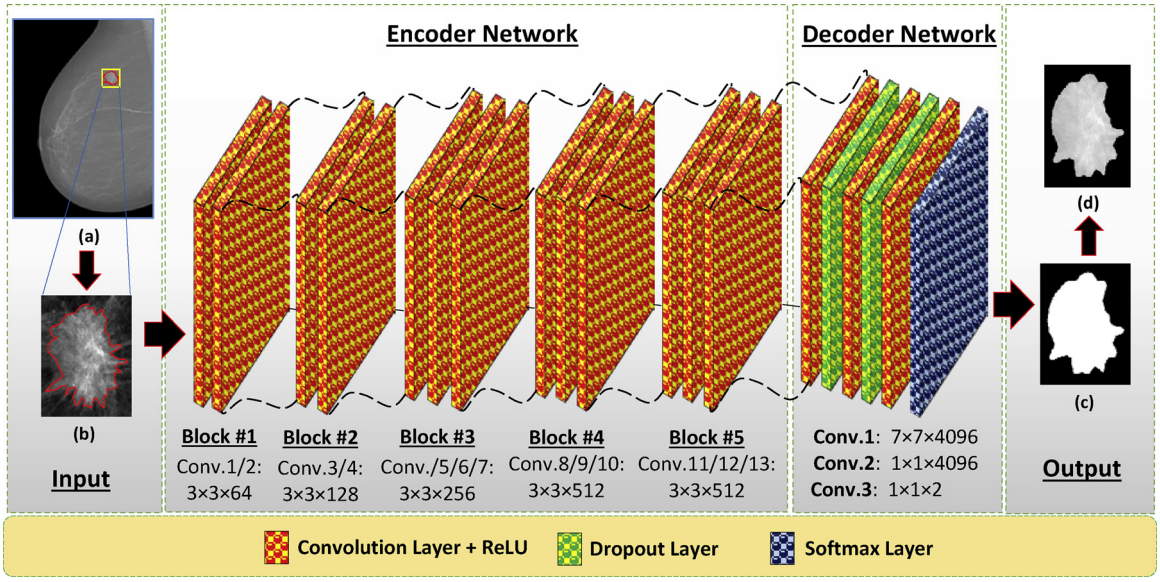

Fig. 2. Proposed deep learning model of a full resolution convolutional network (FrCN) for the mass segmentation stage. (a) Detected ROI (yellow) superimposed on the original mammogram with its ground truth (red), (b) detected ROI (i.e., input mass) with highlighted ground truth (red), (c) output segmented map of input mass, and (d) segmented output mass. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

against other existing deep learning models such as FCN [46], SegNet [48], and U-Net [47] is presented using the same data from the INbreast database [34].

3.4. CNN for mass classification

After detection and segmentation of the masses, a simplified version of AlexNet based on deep CNN (i.e., ConvNet [57]) is used in the classification of the segmented masses as benign or malignant, producing the last output of our CAD system. In general, CNN has a good representative deep model to directly generate deep hierarchical features from the input raw images [19,24,28,33,57,60]. In this study, our CNN consists of five convolutional layers and two FC layers as shown in Fig. 3. For the first and second convolutional layers, 20 and 64 filters with a size of 5 × 5 are used, respectively. The third, fourth, and fifth convolutional layers involve 256 filters with a size of 3 × 3 for each. A non-overlapped max-pooling with a size of 2 × 2 is used to sub-sample the input patch via a factor of 2 as shown in Fig. 3. Meanwhile, local response normalization layers are used after each convolutional layer to improve the performance of the proposed CNN model [48,57]. Then, two FC layers are used with 1024 and 4096 nodes, respectively. After that, a logistic regression layer (i.e., Softmax) with two nodes are added to represent the benign against malignant classification. Finally, the activation function of ReLU is utilized after each stage of CNN except for the last layer which is presented via Softmax. ReLU function is commonly used for deep learning models because its saturation is much faster than sigmoid and tanh functions in terms of training time [57]. Therefore, deep models with ReLU are generally faster and produce better performance as concluded in a previous study [57].

3.5. Experimental settings

In this work, the INbreast database [34] is utilized to evaluate the performance of the proposed CAD system through all three stages, step by step.

For each stage, 4-fold cross-validation tests were carried out with the training, validation, and test datasets, which were generated by stratified partitioning to ensure that each mammogram gets tested equally and to prevent any bias error [10,19,24,29]. This means that all proposed detection, segmentation, and classification deep models are

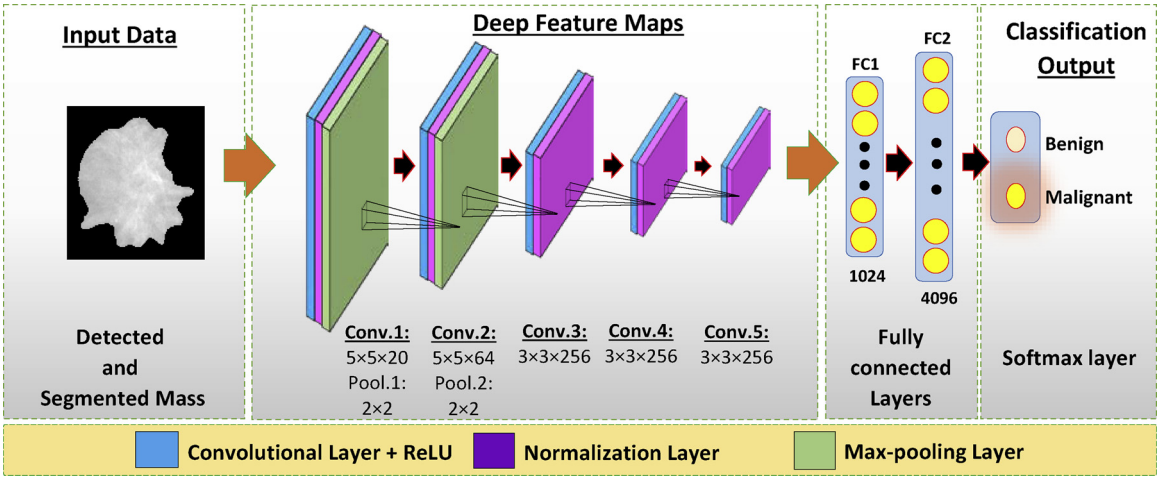

Fig. 3. Proposed deep learning model of the convolutional neural network (CNN) for the mass classification stage. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

Table 1
The performance of the mass detection over 4-fold cross validation via deep learning based YOLO on the test sets of the INbreast dataset.

Fold Test	Benign		Malignant		Total		Metrics (%)		
	True	False	True	False	True	False	Acc.	MCC	F1-score
1st fold	54	0	113	1	167	1	99.40	98.65	99.56
	100%	0.0%	99.12%	0.88%	99.40%	0.59%			
2nd fold	51	3	114	0	165	3	98.21	95.93	98.70
	94.44%	0.05%	100%	0.0%	98.21%	1.79%			
3rd fold	53	1	113	1	166	2	98.81	97.27	99.12
	98.15%	1.85%	99.12%	0.88%	98.81%	1.19%			
4th fold	53	1	114	0	167	1	99.40	98.64	99.56
	98.15%	1.85%	100%	0.0%	99.40%	0.59%			
Average (%)	**97.69**	**0.94**	**99.56**	**0.44**	**98.96**	**1.04**	**98.96**	**97.62**	**99.24**

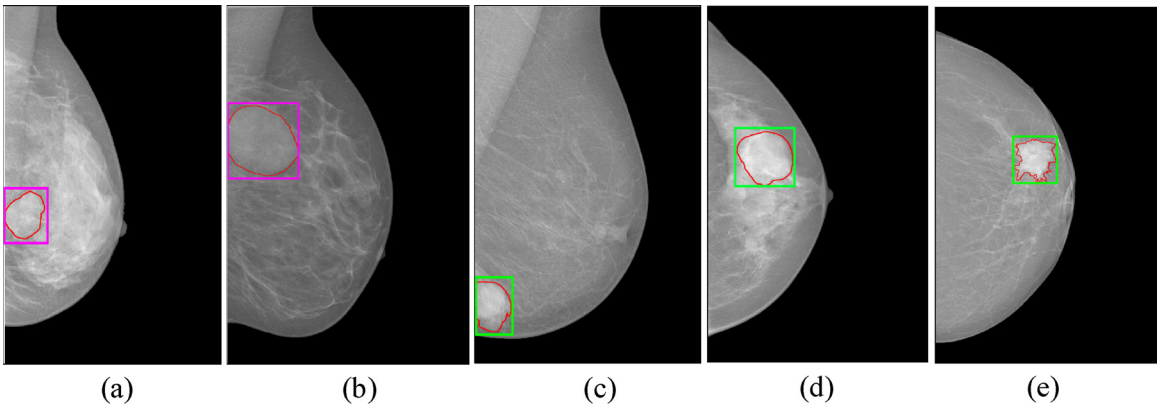

Fig. 4. Examples of mass detection utilizing YOLO on the test images of INbreast dataset. (a) and (b) show the detected ROIs (i.e., masses) for benign cases, while (c), (d), and (e) for malignant cases. Detected ROI are superimposed on the original images: benign (magenta), malignant (green), and ground truth (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

Table 2
Comparison between YOLO-based detection against other methods.

Reference	Method	Data	Testing time per image (Second)	Mass Detection Accuracy (%)
Dhungel et al. [19], Carneiro et al. [24]	Cascade Deep Learning and Random Forest	INbreast	39	96.00
Kozegar et al. [20]	Adaptive threshold with some of machine learning techniques	INbreast	108	87.00
Our method	**Deep learning based YOLO**	**INbreast**	**3**	**98.96**

trained four times to get the overall performance of the proposed CAD system. In all experiments, we randomly divided the augmented dataset for both benign and malignant classes into three groups: 75% (216 benign and 456 malignant) for training, 6.25% (18 benign and 38 malignant) for validation, and 18.75% (54 benign and 114 malignant) for testing as performed in the previous studies [19,29]. To avoid any bias that may occur during the training process due to an unbalance of the training data for all the detection, segmentation, and classification deep learning models, we use the following conditions. First, the training set is shuffled through each mini-batch to make sure each image is utilized only once per epoch as applied by Badrinarayanan et al. [48]. Second, the minimization process utilizing a weighted cross-entropy loss function is used to estimate the parameters of all deep models during the training process as applied in previous studies [24,48,57]. Third, a double cross-validation method is used to select the optimal parameters of all deep learning models with the training and validation datasets. Then, the final performance for all of these models is only assessed utilizing the testing set [19,48,61]. All of these experiments are performed on a PC with the following specifications: Intel (R) Core(TM) i7-6850 K with 16 GB RAM, clock speed or frequency of CPU @ 3.360 GHz, and GPU of NVIDIA GeForce GTX 1080. The CAD system is implemented in Python 2.7.14 and C++ on the Ubuntu 16.04 operating system. The implementation of all deep segmentation models is achieved utilizing Theano [62] and Keras [63] deep learning libraries, while the detection and classification models are implemented under the Tensorflow environment [64].

3.5.1. Experimental settings for mass detection via YOLO

Both views of mammograms (i.e., MLO and CC) in the INbreaset dataset exist with different image sizes [34]. Therefore, all images are resized to 448 × 448 as in previous work [9]. Then, all images are normalized to a range of [0,1]. In this study, we consider the masses to be correctly detected if the intersection over union ($IoU^{Extracted}_{Ground\ truth}$) between the extracted and annotated (i.e., ground truth) bounding boxes of the mass is greater than or equals 50% as in previous work [4,9,10,19,20]. Moreover, the false positive candidates of ROIs are manually excluded before the segmentation and classification stages of the CAD system as applied previously [19,24]. This is because there is a lack of ground truth for the detected false masses to derive the performance evaluation metrics [19]. Thus, the evaluation process for the segmentation and classification stages is done with the exception of the falsely detected cases.

3.5.2. Experimental settings for mass segmentation via FrCN

In the segmentation stage, only correctly detected masses are used, while the falsely detected are manually excluded as previously done in

Table 3
Segmentation performance of our proposed full resolution convolutional network (FrCN) against other methods on the test sets of the INbreast dataset.

Test Fold	Method	Measurement metrics (%)						
		Dice	Jac.	Sen.	Spe.	Acc.	AUC	MCC
1st fold	FCN	89.11	80.36	82.97	96.98	90.23	89.98	81.07
	SegNet	89.32	80.7	82.75	**97.64**	90.47	90.19	81.64
	U-Net	90.26	82.25	84.6	97.34	91.2	90.97	82.94
	Proposed FrCN	**92.75**	**86.47**	**92.07**	93.99	**93.06**	**93.03**	**86.11**
2nd fold	FCN	88.81	79.87	82.50	96.99	90.06	89.74	80.73
	SegNet	88.88	79.99	82.35	97.31	90.16	89.83	80.97
	U-Net	89.82	81.51	83.71	97.53	90.93	90.62	82.43
	Proposed FrCN	**92.75**	**86.48**	**92.98**	93.12	**93.05**	**93.05**	**86.09**
3rd fold	FCN	89.15	80.42	83.18	96.76	90.15	89.97	80.92
	SegNet	88.69	79.68	81.65	97.67	89.88	89.66	80.63
	U-Net	89.42	80.87	82.98	97.52	90.45	90.25	81.63
	Proposed FrCN	**92.89**	**86.72**	**92.90**	93.26	**93.09**	**93.08**	**86.16**
4th fold	FCN	88.54	79.44	82.26	96.78	89.87	89.52	80.31
	SegNet	89.40	80.83	83.31	97.22	90.60	90.26	81.74
	U-Net	91.17	83.77	87.73	95.70	91.91	91.72	83.95
	Proposed FrCN	**92.36**	**85.81**	**92.94**	92.47	**92.69**	**92.70**	**85.36**
Average	FCN	88.90	80.02	82.72	96.88	90.08	89.80	80.76
	SegNet	89.07	80.30	82.51	97.46	90.28	89.99	81.25
	U-Net	90.17	82.10	84.76	97.02	91.12	90.89	82.74
	Proposed FrCN	**92.69**	**86.37**	**92.72**	93.21	**92.97**	**92.97**	**85.93**

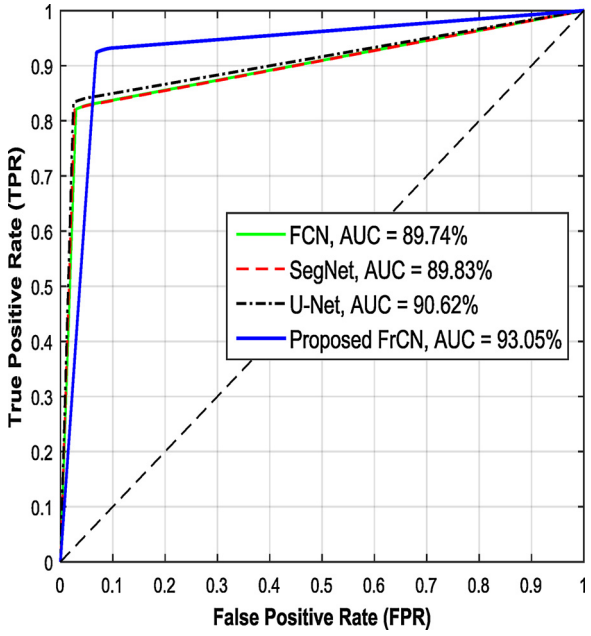

Fig. 5. The performance of mass segmentation in terms of ROC curves on the test sets of INbreast dataset.

[19,24]. Similar to the detection stage, the same 4-fold cross validation is performed for all segmentation deep learning models: FCN [46], SegNet [48], U-Net [47], and the proposed FrCN. To train all of these deep learning models, a learning rate of 0.001 with Adam optimizer is utilized [52,65]. Meanwhile, 100 epochs and 20 mini-batches are used to optimize and select the model parameters with the training and validation datasets. A weighted cross-entropy is used as a loss function to handle the variety of the pixel numbers in each class during the training process. As shown in Fig. 2, a dropout of 0.5 is added after the first and second convolutional layers in the decoder network to prevent the overfitting.

3.5.3. Experimental settings for mass classification via CNN

All segmented masses are resized to a size of 40 × 40 utilizing a bicubic interpolation [19,21,24,33,45,57]. These masses are then fed into the classification stage via the implemented CNN as shown in Fig. 3. In order to show the impact of the segmentation, we have tested the CAD systems without and with the segmentation stage. That is, in *the former*, the correctly detected masses are directly fed into the classification via CNN ignoring the segmentation stage. In *the latter*, the detected masses are first segmented via the proposed FrCN and then passed into the classification stage of our CNN. This comparison is performed under the same settings and architectures of the classifier for CAD systems. To verify both settings of the CAD systems during classification, the same 4-fold cross validation is performed similar to in the detection and segmentation stages. For training, an Adam optimizer with the learning rate of 0.0001 and weight decay of 0.0005 is utilized [52,65]. Meanwhile, the number of epochs and mini-batch size are set to 100 and 24, respectively. A dropout of 0.3 is used on both fully connected layers to accelerate the training process as well as prevent the overfitting [19,29,48,66].

3.6. Evaluation metrics

The three stages of the proposed CAD system are evaluated separately as follows. Overall accuracy, is used to evaluate the detection stage. Since the dataset used in this study is unbalanced, we also used F1-score and Matthews correlation coefficient (MCC). The F1-score is also known as the Dice similarity coefficient which represents a harmonic average of the precision and sensitivity. Its maximum score of 1 indicates perfect precision and sensitivity and of 0 the worst [25]. MCC is used as a balanced measure of the quality of classifications even if the classes are in different sizes, counting in true and false positives and negatives [67]. Sensitivity, specificity, overall accuracy, Dice similarity coefficient or F1-score, Jaccard index, and MCC are used to evaluate the proposed segmentation method of FrCN against others (i.e., FCN, SegNet, and U-Net) [19,52,68,25]. The criteria for all of these metrics are defined as follows,

$$\text{Sensitivity (Sen.)} = \frac{TP}{TP + FN}, \tag{1}$$

$$\text{Specificity (Spe.)} = \frac{TN}{TN + FP}, \tag{2}$$

$$\text{Jaccard (Jac.)} = \frac{TP}{TP + FP + FN}, \tag{3}$$

$$\text{F1}-\text{score (Dice)} = \frac{2 \bullet TP}{2 \bullet TP + FP + FN}, \tag{4}$$

$$\text{Overall accuracy (Acc.)} = \frac{TP + TN}{TP + FN + TN + FP}, \tag{5}$$

$$MCC = \frac{TP \bullet TN - FP \bullet FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{6}$$

where TP, TN, FP, and FN are defined per pixel to represent the number of true positive, true negative, false positive, and false negative detections, respectively. The confusion matrix is used to derive all of these parameters. A good performance of segmentation is achieved with high sensitivity and specificity where all the masses and surrounding tissues are correctly segmented. Meanwhile, the rate of similarity between predicted and ground-truth regions is measured using the Dice (F1-score) and Jaccard metrics interpreting how many TP pixels are found to be FPs [19,25]. The Matthew correlation coefficient (MCC) is also used to measure the correlation between the segmented mass pixels and its ground-truth [69]. Moreover, a tradeoff between sensitivity and

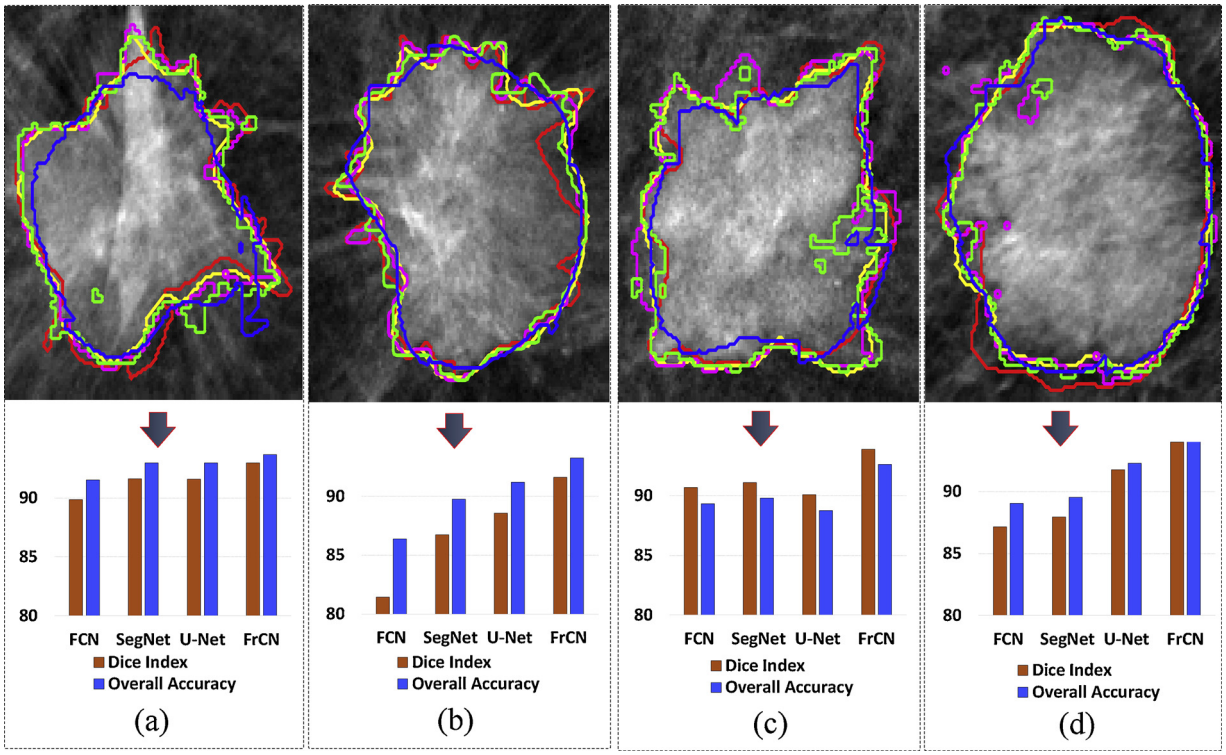

Fig. 6. Examples of the segmentation results for the proposed full resolution convolutional network (FrCN) against the fully convolutional network (FCN), U-Net, and SegNet. The contours indicate the ground truth (red), FrCN (yellow), U-Net (green), SegNet (magenta), and FCN (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

Table 4
Comparison of segmentation time for the full resolution convolutional network (FrCN) against other methods.

Method	Training time per epoch (Second)	Testing time per image (Second)
FCN	466	10.25
SegNet	245	10.48
U-Net	289	10.66
Proposed FrCN	**231**	**8.51**

specificity, producing ROC curves with AUC, is also used to evaluate the segmentation method [70]. For classification, sensitivity, specificity, overall accuracy, ROC curve with AUC, MCC and F1-score are used per image (i.e., input ROI) not per pixel as in the segmentation stage [3,4,9,19,29,33].

4. Results

4.1. Mass detection results

The performance of mass detection over the four-fold cross-validation on the INbreast test datasets is reported in Table 1. Examples of the mass detection results showing the potential ROIs (masses) are shown in Fig. 4. In each test fold, the detected regions are considered to be correct when IoU $\geq$ 50%. The false detections presented in Table 1 indicate those cases with IoU < 50%. An average of 98.96% in overall detection accuracy, MCC of 97.62% and F1-score of 99.24% demonstrate the reliable performance of the YOLO detector. A comparison of results using YOLO against the latest methods are listed in Table 2. It is clearly observed that the YOLO detector performed better in accuracy. In addition, it is much faster than other deep models as presented in Table 2.

4.2. Mass segmentation results

As shown in Table 1, the falsely detected cases of breast masses over each test fold have been excluded in the segmentation stage. The segmentation performances of the proposed FrCN against FCN (i.e., FCN-8 [46]), SegNet, and U-Net are presented in Table 3. These results are measured from each fold of testing in the same set as presented in the detection stage. Here, all quantitative measurements for the mass segmentation are computed per pixel of the segmented maps with the same resolution of the original image (i.e., input ROI).

As shown in Table 3, FrCN clearly outperforms other methods with an average Dice index of 92.69%, Jaccard coefficient of 86.37%, overall

Table 5
Comparison of classification performance (%) for both CAD systems over 4-fold cross validation on the test sets of the INbreast dataset.

Fold Test	CAD system without mass segmentation						CAD system with mass segmentation					
	Sen.	Spe.	Acc.	AUC	MCC	F1-score	Sen.	Spe.	Acc.	AUC	MCC	F1-score
1st fold	92.04	90.74	91.62	91.39	81.32	93.69	98.23	92.59	96.41	95.41	91.74	97.47
2nd fold	92.98	86.27	90.91	89.63	78.84	93.39	97.37	92.16	95.76	94.76	90.02	96.94
3rd fold	92.04	88.68	90.96	90.36	79.59	93.27	97.35	92.45	95.78	94.90	90.26	96.92
4th fold	92.11	88.68	91.02	90.39	79.65	93.33	95.61	92.45	94.61	94.03	87.63	96.04
Average	**92.29**	**88.59**	**91.13**	**90.61**	**79.85**	**93.42**	**97.14**	**92.41**	**95.64**	**94.78**	**89.91**	**96.84**

Table 6
Confusion matrices of classification via a convolutional neural network (CNN) with and without mass segmentation over 4-fold cross validation on the test sets of the INbreast dataset.

Fold Test	CAD system without mass segmentation			CAD system with mass segmentation	
	Actual Classes	*Predicted Classes*		*Predicted Classes*	
		Benign	Malignant	Benign	Malignant
1st fold	Benign	49	5	50	4
		90.74%	9.23%	92.59%	7.41%
	Malignant	9	104	2	111
		7.96%	92.04%	1.77%	98.23%
2nd fold	Benign	44	7	47	4
		86.27%	13.73%	92.16%	7.84%
	Malignant	8	106	3	111
		7.02%	92.98%	2.63%	97.34%
3rd fold	Benign	47	6	49	4
		88.68%	11.32%	92.45%	7.55%
	Malignant	9	104	3	110
		7.96%	92.04%	2.65%	97.35%
4th fold	Benign	47	6	49	4
		88.68%	11.32%	92.45%	7.55%
	Malignant	9	105	5	109
		7.89%	92.11%	4.39%	95.61%
Average (%)	Benign	88.59	11.40	92.41	7.59
	Malignant	7.71	92.29	2.86	97.13

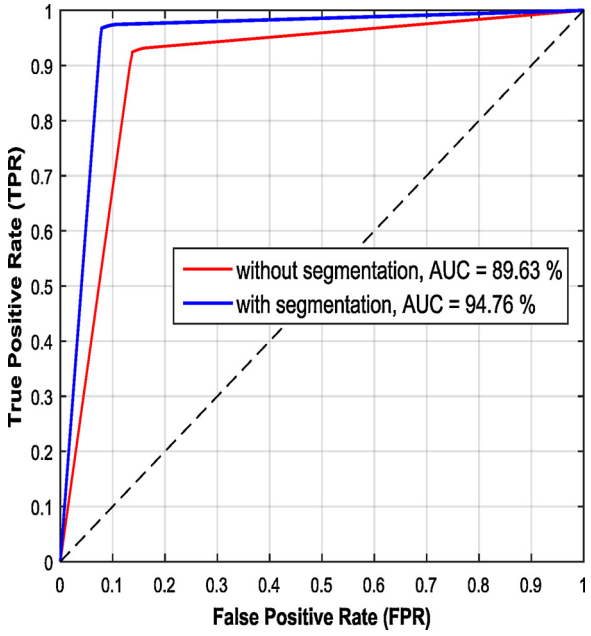

Fig. 7. The performance of mass classification for both schemes of CAD system (i.e., with and without segmentation) in terms of ROC curves on the test sets of the INbreast dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

accuracy of 92.97%, and MCC of 85.93%. In contrast, SegNet achieves a better specificity performance of 97.46%. Moreover, the segmentation performance of FrCN against all other methods was evaluated by the AUC over all test folds. Fig. 5 shows an example of the ROC curves with their AUCs from the 2nd test fold for comparison among all methods. As clearly shown in Table 3 and Fig. 5, the performance of FrCN outperforms all other methods with an average AUC of 92.97%. Examples of the qualitative segmentation results for the proposed FrCN against FCN, SegNet, and U-Net are shown in Fig. 6. Additionally, a comparison of Dice index and overall accuracy for each corresponding image (mass) are presented. Table 4 shows the comparison of computing time for FrCN against FCN, SegNet, and U-Net for training and testing durations. As presented in Tables 3 and 4, FrCN achieves better mass segmentation results and faster training time (6.42 h) for all epochs than FCN, SegNet, and U-Net with 12.94, 6.81, and 8.03 h, respectively. Although U-Net achieves better segmentation results than SegNet, it is slower in the training and testing tasks.

4.3. Mass classification results

All segmented masses via the proposed FrCN are sequentially fed into the classification stage in the same test folds from the previous detection and segmentation stages. The performance of classification is evaluated in terms of sensitivity, specificity, overall accuracy, ROC curve with its AUC, MCC, and F1-score values for the CAD systems with and without the segmentation stage as reported in Table 5. It is obviously noted that the CAD system with the segmentation achieves much better results with a sensitivity of 97.14%, specificity of 92.41%, overall accuracy of 95.64%, and AUC of 94.78% as well as MCC and F1-score are improved by 10.06% and 3.42%, respectively. Meanwhile, the confusion metrics with each test fold for both cases of the CAD systems (i.e., with and without segmentation) are presented in Table 6. For the CAD system with segmentation, it is obviously demonstrated that 92.41% of benign and 97.13% of malignant cases are correctly classified, while 2.86% of benign and 7.59% of malignant are negatively classified. However, the false positive rates are negatively affected by the specificity through both cases of the CAD system (i.e., with and without the segmentation) by 11.40% and 7.59%, respectively. Thereby, the CAD system with the segmentation provided much better results over all 4-fold cross validations in all measurements as presented in Tables 5 and 6. The results shown in Tables 5 and 6 indicate a stable performance by the CAD system through all fold subsets. The results of classification show the robustness of the proposed CAD system, minimizing the false positive and negative rates. An example of ROC curves with AUCs for both cases of the CAD system on the 2nd test fold is shown in Fig. 7. As shown in Fig. 7, the TPR is increased and FPR is decreased which indicates that the CAD system performs better.

5. Discussion

Deep learning based CNNs have recently achieved remarkable success in the analysis of medical images [25,54,55], In this study, a fully integrated CAD system based on deep learning covering detection,

Table 7
Comparison between the performances of the proposed computer-aided diagnosis (CAD) system based on deep learning through all stages of detection, segmentation, and classification against others in the latest studies on the test sets.

Reference	Data	Prediction classes	Total testing time per image (Second)	Overall accuracy/ (AUC) (%)	Hardware specs
Dhungel et al. [19]	INbreast	Benign/Malignant	41	91 / (76)	Intel Core i5-2500k, 8GB RAM, 3.30 GHz, and GPU of NVIDIA GeForce GTX 460 SE 4045 MB
Carneiro et al. [24]	INbreast	Normal/Benign/ Malignant	NA	NA / (78: *Benign Vs. Malignant*) & NA / (86: *Malignant Vs. Normal + Benign*)	Intel Core i7, 8GB RAM, 2.3 GHz, and GPU of NVIDIA GeForce GT 650 M 1024 MB
Proposed CAD system	**INbreast**	**Benign/Malignant**	**12.23**	**95.64 / (94.78)**	**Intel Core i7-6850 K, 16 GB RAM, 3.360 GHz, and GPU of NVIDIA GeForce GTX 1080**

segmentation, and classification stages are presented. Recent studies involving CAD systems showed that mass detection encounters challenges, especially when the masses exist inside dense tissues or over pectoral muscles in the breast images [4,19,24]. YOLO-based deep learning can detect the masses even if they exist inside dense tissues or over pectoral muscles as shown in Fig. 4(a) and (b), respectively. The proposed YOLO detector plays a critical role in the CAD system, achieving the best detection performance compared with the latest deep learning models [14,19,24].

In order to overcome the mass segmentation challenge as well as achieve a better classification using a CAD system, a new method of FrCN for pixel-to-pixel mass segmentation is proposed. In fact, segmentation generates more specific and representative shape features conducted with the mass regions improving the final classification results [19,21,25]. The proposed FrCN overcomes the limitations of the latest deep learning segmentation models in terms of preserving high resolution the details. As concluded by Yu et al. in 2017, a much better classification performance can be achieved when segmentation of a skin lesion from dermoscopy images is outperformed [25]. For comparison, the results of all methods presented in Table 3 and Fig. 6 are achieved without any refinement pre- and/or post-processing methods. Moreover, the segmentation of the mass improves the classification rates of the proposed CAD system. For mass segmentation, each pixel (i.e., pixel-to-pixel) in the input image (i.e., detected mass) has its own label making the training process of the proposed FrCN less complex. This means each pixel represents an independent sample during the training process which highly increases the number of training samples as well.

For mass classification via CNN, the proposed CAD system achieves much better results when mass segmentation is utilized, with an overall accuracy of 95.64%, MCC of 89.91% and F1-score of 96.84%, against 91.13%, 79.85%, and 93.42% without mass segmentation, respectively. The improved performance is due to the following reasons. First, the suspicious regions of the masses are accurately detected via YOLO. Second, the robust segmentation based on the deep learning FrCN model plays a crucial role in segmenting the specific region of masses excluding the surrounding tissues and then the amount of false positive and negative pixels is decreased. Third, utilizing the high deep level features from the proposed deep learning model of CNN contributed to improving the performance of the proposed CAD system. This means that an accurate detection and segmentation of masses is important to achieve a more feasible and reliable CAD system. Finally, a comparison between our proposed CAD system through our methodology with respect to the latest work in the field is presented in Table 7. Our proposed CAD system could handle all stages of detection, segmentation, and classification with a higher performance and in a faster time than others with a total testing time for all stages of 12.23 s as summarized in Table 7. Therefore, the performance of the proposed CAD system seems to make its practical application possible.

6. Conclusion

In this paper, a fully integrated CAD system based on deep learning for detection, segmentation, and classification of the masses from mammograms in a single framework is presented. The development of a practical CAD system, which automatically detects masses, segments and predicts their types either benign or malignant is required. To detect masses from an entire mammogram in the most challenging cases, YOLO-based on deep learning can be used and outperforms other methods. Due to the segmentation capability of the proposed deep model using a FrCN, the CAD system can achieve a much better classification results. Hence, the mass segmentation pixel-to-pixel could be a key to decrease the false positive and negative rates of pixels and then improve the overall performance of the proposed CAD system. Classification results of CNN based on the segmentation presents the efficiency and feasibility of the proposed CAD system compared to the latest methods in the field. The proposed CAD system based deep learning through detection, segmentation, and classification could be used for clinical applications to assist radiologists.

Conflict of interest

All authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by International Collaborative Research and Development Programme funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) (N0002252).

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2017, CA. Cancer J. Clin. 67 (no. 1) (2017) 7–30.
[2] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, M. Thun, Cancer statistics, 2008, A Cancer J. Clin. 58 (no. 2) (2008) 71–96.
[3] M.A. Al-antari, M.A. Al-masni, S.U. Park, J.H. Park, Y.M. Kadah, S.M. Han, T.-S. Kim, Automatic computer-aided diagnosis of breast cancer in digital mammograms via deep belief network, Global Conference on Engineering and Applied Science (GCEAS), Japan, 2016, pp. 1306–1314.
[4] M.A. Al-antari, M.A. Al-masni, S.U. Park, J.H. Park, M.K. Metwally, Y.M. Kadah, S.M. Han, T.-S. Kim, An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network, J. Med. Biol. Eng. 38 (no. 3) (2017) 443–456, http://dx.doi.org/10.1007/s40846-017-0321-6.
[5] R. Takahashi, Y. Kajikawa, Computer-aided diagnosis: a survey with bibliometric analysis, Int. J. Med. Inf. 101 (2017) 58–67.
[6] Y. Wang, D. Tao, X. Gao, X. Li, B. Wang, Mammographic mass segmentation: embedding multiple features in vector-valued level set in ambiguous regions, Pattern Recognit. 44 (no. 9) (2011) 1903–1915.
[7] S. Lee, C. Lo, C. Wang, P. Chung, C. Chang, C. Yang, P. Hsu, A computer-aided design mammography screening system for detection and classification of microcalcifications, Int. J. Med. Inf. 60 (no. 1) (2000) 29–57.
[8] A. Casellas-Grau, J. Vives, A. Font, C. Ochoa, Positive psychological functioning in breast cancer: an integrative review, Breast 27 (2016) 136–168.
[9] M. Al-masni, M. Al-antari, j. Park, G. Gi, T. Kim, P. Rivera, E. Valarezo, S.-M. Han, T.-s. Kim, Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network, 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'17), Jeju Island, South Korea, 2017, pp. 1230–1236.
[10] M.A. Al-masni, M. Al-antari, J.-m.P. Park, G. Gi, T.-Y.K. Kim, P. Rivera, E. Valarezo, M.-T. Choi, S.-M. Han, T.-S. Kim, Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system, Comput. Methods Prog. Biomed. 157 (2018) 85–94.
[11] M.A. Al-antari, M.A. Al-masni, Y.M. Kadah, Hybrid model of computer-aided breast cancer diagnosis from digital mammograms, J. Sci. Eng. 04 (no. 2) (2017) 114–126.
[12] M.L. Giger, Medical imaging and computers in the diagnosis of breast cancer, SPIE Opt. Eng. Appl. (2014) 918908–918908.
[13] C. Dromain, B. Boyer, R. Ferré, S. Canale, S. Delaloge, C. Balleyguier, Computed-aided diagnosis (CAD) in the detection of breast cancer, Eur. J. Radiol. 82 (no. 3) (2013) 417–423.
[14] N. Dhungel, G. Carneiro, A.P. Bradley, Automated mass detection in mammograms using cascaded deep learning and random forests, International Conference on Digital Image Computing: Techniques and Applications (DICTA), Australia, 2015, 2015.
[15] C. Muramatsu, T. Hara, T. Endo, H. Fujita, Breast mass classification on mammograms using radial local ternary patterns, Comput. Biol. Med. 72 (no. 1) (2016) 43–53.
[16] J. Virmani, N. Dey, V. Kumar, PCA-PNN and PCA-SVM Based CAD Systems for Breast Density Classification, Warsaw, Springer International Publishing, Poland, 2016, pp. 159–180.
[17] N.I. Yassin, S. Omran, E.M. Houby, H. Allam, Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review, Comput. Methods Prog. Biomed. 156 (2018) 25–45.
[18] H. Li, X. Meng, T. Wang, Y. Tang, Y. Yin, Breast masses in mammography classification with local contour features, Biomed. Eng. Online 16 (no. 1) (2017) 44–54.
[19] N. Dhungel, G. Carneiro, A.P. Bradley, A deep learning approach for the analysis of masses in mammograms with minimal user intervention, Med. Image Anal. 37 (no. 1) (2017) 114–128.
[20] E. Kozegar, M. Soryani, B. Minaei, D. Inês, Assessment of a novel mass detection algorithm in mammograms, J. Cancer Res. Ther. (2013) 592–600.
[21] N. Dhungel, G. Carneiro, A.P. Bradley, Deep structured learning for mass segmentation from mammograms, Image Processing (ICIP), 2015 IEEE Inter- National Conference, (2015), pp. 2950–2954.
[22] J. Wei, B. Sahiner, L.M. Hadjiiski, H.-P. Chan, N. Petrick, M. Helvie, M.A. Roubidoux, J. Ge, Computer-aided detection of breast masses on full field digital mammograms, Med. Phys. 32 (no. 9) (2005) 2827–2838.
[23] L. Oliveira, A. Silva, L. VilelaRibeiro, R. Oliveira, C. Coelho, A. Andradea,

Computer-aided diagnosis in chest radiography for detection of childhood pneumonia, Int. J. Med. Inf. 77 (no. 8) (2008) 555–564.
[24] G. Carneiro, J. Nascimento, A.P. Bradley, Automated analysis of unregistered multi-view mammograms with deep learning, IEEE Trans. Med. Imaging 36 (no. 11) (2017) 2355–2365.
[25] L. Yu, H. Chen, Q. Dou, J. Qin, P.-A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE Trans. Med. Imaging 36 (no. 4) (2017) 994–1004.
[26] P. Rahmati, A. Adler, G. Hamarneh, Mammography segmentation with maximum likelihood active contours, Med. Image Anal. 16 (no. 9) (2012) 1167–1186.
[27] A.R. Domínguez, A. Nandi, Toward breast cancer diagnosis based on automated segmentation of masses in mammograms, Pattern Recognit. 42 (no. 6) (2009) 1138–1148.
[28] Y. Qiu, S. Yan, R.R. Gundreddy, Y. Wang, S. Cheng, H. Liu, B. Zheng, A New approach to develop computer-aided diagnosis Scheme of breast mass classification using deep learning technology, J. X-Ray Sci. Technol. 25 (no. 5) (2017) 751–763.
[29] T. Kooi, G. Litjens, Bv. Ginneken, A.M. Gubern, C.I. Sánchez, R. Mann, Ad. Heeten, N. Karssemeijer, Large scale deep learning for computer aided detection of mammographic lesions, Med. Image Anal. 35 (2017) 303–312.
[30] N. Dhungel, G. Carneiro, A.P. Bradley, DEEP STRUCTURED LEARNING FOR MASS SEGMENTATION FROM MAMMOGRAMS, Image Processing (ICIP), 2015 IEEE Inter- National Conference, (2015), pp. 2950–2954.
[31] C. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag, New York, USA, 2006.
[32] J. Chakraborty, A. Midya, R. Rabidas, Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns, Expert Syst. Appl. 99 (no. 1) (2018) 168–179.
[33] Z. Jiao, X. Gao, Y. Wang, J. Li, A deep feature based framework for breast masses classification, Neurocomputing 197 (no. C) (2016) 221–231.
[34] I. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. Cardoso, J. Cardoso, INbreast: toward a full-field digital mammographic database, Acad. Radiol. 19 (no. 2) (2012) 236–248.
[35] A.-B. Ayelet, L. Karlinsky, S. Alpert, S. Hasoul, R. Ben-Ari, E. Barkan, A region based convolutional network for tumor detection and classification in breast mammography, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Athens, Greece, Springer International Publishing, 2016, pp. 197–205 pp. 197-205.
[36] M. Heath, K. Bowyer, D. Kopans, R. Moore, W.P. Kegelmeyer, The digital database for screening mammography, 5th International Workshop on Digital Mammography, (2001).
[37] J. Arevalo, F.A. González, R. Ramos-Pollán, J.L. Oliveira, M.A.G. Lopezd, Representation learning for mammography mass lesion classification with convolutional neural networks, Comput. Methods Prog. Biomed. 127 (2016) 248–257.
[38] T. Kooi, A. Gubern-Merida, J.-J. Mordang, R. Mann, R. Pijnappel, K. Schuur, Ad. Heeten, N. Karssemeije, A comparison between a deep convolutional neural network and radiologists for classifying regions of interest in mammography, International Workshop on Digital Mammography, Sweden, 2016, pp. 51–56.
[39] Y. Zhang, K. Sohn, R. Villegas, G. Pan, H. Lee, Improving object detection with deep convolutional networks via bayesian optimization and structured prediction, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2015), pp. 249–258.
[40] M. Kupinski, M. Giger, Automated seeded lesion segmentation on digital mammograms, IEEE Trans. Med. Imaging 17 (no. 4) (1998) 510–517.
[41] J.S. Cardoso, I. Domingues, H.P. Oliveira, Closed shortest path in the origi- nal coordinates with an application to breast cancer, Int. J. Pattern Recognit. Artif. Intell. 29 (no. 1) (2015) 2.
[42] M. Xiao, S. Xia, S. Wang, Geometricactivecontourmodelwithcolorand intensity priorsformedicalimagesegmentation, Proceedings of the 27th Annual International Conference of the IEEE Engineeringin Medicineand Biology, (2005), pp. 6496–6499.
[43] Y. Yuan, M. Giger, H. Li, K. Suzuki, C. Sennett, Adual-stagemethodforlesion segmentation ondigitalmammograms, Med. Phys. 34 (no. 11) (2007) 4180–4193.
[44] M. Yu, Q. Huang, J. Song, E. Liu, H. Hung, A novel segmentation method for convex lesions based on dynamic programming with local intra–class variance, 27th Annual ACM Symposium on Applied Computing, (2012), pp. 39–44.
[45] N. Dhungel, G. Carneiro, A.P. Bradley, Deep learning and structured prediction for the segmentation of mass in mammograms, International Conference on Medical Image Computing and Computer-Assisted Intervention, (2015), pp. 605–612.
[46] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (no. 4) (2017) 640–651.
[47] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, (2015).
[48] V. Badrinarayanan, A. Kendall and R. Cipoll, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, in arXiv preprint arXiv:1511.00561, 2016.
[49] H. Noh, S. Hong and B. Han, Learning Deconvolution Network for Semantic Segmentation," in arXiv:1505.04366v1, 2015.
[50] L.-C. Chen, G. Papandreou and I. Kokki, DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," in arXivpreprint arXiv:1606.00915, 2017.
[51] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in arXiv preprintarXiv:1409.1556, 2014.
[52] Y. Yuan, M. Chao, Y.-C. Lo, Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance, IEEE Trans. Med. Imaging 36 (no. 9) (2017) 1876–1886.
[53] A. Jorstad, P. Fua, Refining mitochondria segmentation in electron mi- croscopy imagery with active surfaces, Computer Vision-ECCV 2014 Work- Shops, (2014), pp. 367–379.
[54] G. Litjens, T. Kooi, B.E. Bejnordi, A. Setio, F. Ciompi, M. Ghafoorian, J. Laak, B. Ginneken, C. anchez, A survey on deep learning in medical image analysis, arXiv:1702.05747v2 [Cs.CV], (2017) 4 Jun 2017.
[55] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, J. Garcia-Rodriguez, A review on deep learning techniques applied to semantic segmentation, arXiv:1704.06857 [Cs.CV], (2017).
[56] R. Llobet, J. Perez-Cortes, A. Toselli, A. Juan, Computer-aided detection of prostate cancer, Int. J. Med. Inf. 76 (no. 7) (2007) 547–556.
[57] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, 25th International Conference on Neural Information Processing Systems, USA, 2012, pp. 1097–1105.
[58] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, IEEE Conference on Computer Vision and Pattern Recognition, (2016).
[59] J. Ball, L. Bruce, Digital mammographic computer aided diagnosis (CAD) using adaptive level set segmentation, 29th Annual International Conference of the IEEE (EMBS), (2007), pp. 4973–4978.
[60] J.H. Park, S.U. Park, M. Zia Uddin, M. Al-antari, M. Al-masni, T.-S. Kim, A single depth sensor based human activity recognition via convolutional neural network, 4th World Conference on Applied Sciences, Engineering & Technology, (2016), pp. 329–332.
[61] E. Szymańska, E. Saccenti, A.K. Smilde, J.A. Westerhuis, Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies, Metabolomics 8 (no. 1) (2012) 3–16.
[62] L. lab, Theano, [Online] Available: University of Montreal, 2017 (Accessed 10, 2017), http://deeplearning.net/software/theano/tutorial/.
[63] F. Chollet, Keras: The Python Deep Learning Library, [Online]. Available: MIT, 2017 (Accessed 10, 2017), https://keras.io/.
[64] Google Brain Team, TensorFlow, 9 11 2017. [Online]. Available: (2017) (Accessed 10, 2017), www.tensorflow.org.
[65] K.P. Diederik, J.L. Ba, Adam: a method for stochastic optimization, the 3rd International Conference for Learning Representations, San Diego, 2015, pp. 1–15.
[66] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple Way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.
[67] G. Jurman, S. Riccadonna, C. Furl, A comparison of MCC and CEN error measures in multi-class prediction, PLoSE ONE (2012).
[68] M. Goyal, M.H. Yap, Multi-class semantic segmentation of skin lesions via fully convolutional networks, arXiv:1711.10449 [Cs.CV], (2017).
[69] D.M. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, J. Mach. Learn. Technol. 2 (no. 1) (2011) 37–63.
[70] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, D. Feng, Dermoscopic image segmentation via multistage fully convolutional networks, IEEE Trans. Biomed. Eng. 64 (no. 9) (2017) 2065–2074.