Deep Learning Computer-Aided Diagnosis for Breast Lesion in Digital Mammogram

Mugahed A. Al-antari, Mohammed A. Al-masni, and Tae-Seong Kim

Abstract

For computer-aided diagnosis (CAD), detection, segmentation, and classification from medical imagery are three key components to efficiently assist physicians for accurate diagnosis. In this chapter, a completely integrated CAD system based on deep learning is presented to diagnose breast lesions from digital X-ray mammograms involving detection, segmentation, and classification. To automatically detect breast lesions from mammograms, a regional deep learning approach called You-Only-Look-Once (YOLO) is used. To segment breast lesions, full resolution convolutional network (FrCN), a novel segmentation model of deep network, is implemented and used. Finally, three conventional deep learning models including regular feedforward CNN, ResNet-50, and InceptionResNet-V2 are separately adopted and used to classify or recognize the detected and segmented breast lesion as either benign or malignant. To evaluate the integrated CAD system for detection, segmentation, and classification, the publicly available and annotated INbreast database is used over fivefold cross-validation tests. The evaluation results of the YOLO-based detection achieved detection accuracy of 97.27%, Matthews's correlation coefficient (MCC) of 93.93%, and F1-score of 98.02%. Moreover, the results of the breast lesion segmentation via FrCN achieved an overall accuracy of 92.97%, MCC of 85.93%, Dice (F1-score) of 92.69%, and Jaccard similarity coefficient of 86.37%. The detected and segmented breast lesions are classified via CNN, ResNet-50, and InceptionResNet-V2 achieving an average overall accuracies of 88.74%, 92.56%, and 95.32%, respectively. The performance evaluation results through all stages of detection, segmentation, and classification show that the integrated CAD system outperforms the latest conventional deep learning methodologies. We conclude that our CAD system could be used to assist

Mugahed A. Al-antari and Mohammed A. Al-masni contributed equally with all other contributors.

M. A. Al-antari
Department of Biomedical Engineering, College of Electronics and Information, Kyung Hee University, Yongin, Republic of Korea

Department of Biomedical Engineering, Sana'a Community College, Sana'a, Republic of Yemen

M. A. Al-masni · T.-S. Kim (✉)
Department of Biomedical Engineering, College of Electronics and Information, Kyung Hee University, Yongin, Republic of Korea
e-mail: tskim@khu.ac.kr

G. Lee, H. Fujita (eds.), *Deep Learning in Medical Image Analysis*, Advances in Experimental Medicine and Biology 1213,
https://doi.org/10.1007/978-3-030-33128-3_4

radiologists over all stages of detection, segmentation, and classification for diagnosis of breast lesions.

Keywords
Medical image analysis · Mammograms · Breast lesion · Computer-aided diagnosis (CAD) · Deep learning · Full resolution convolutional network (FrCN) · Detection · Segmentation · Classification

Introduction

Breast cancer is the second most common cancer affecting the life of women worldwide [1]. In 2019, breast cancer is statistically categorized among the highest causes of all other cancers, accounting for 30% of estimated new cases and 15% of death cases [1]. Early detection of breast cancer is a crucial requirement to reduce the mortality rate among women [1]. Up to date, digital X-ray mammography is the standard and most reliable tool to screen out the suspicious breast masses and microcalcifications for patients [2]. In 2015, the American Cancer Society (ACS) updated its breast cancer screening roles. Currently, ACS recommends women over 45 years to undergo breast screening one time per year using double views of mammograms: Mediolateral Oblique (MLO) and Cranio-Caudal (CC). Whereas, women with age of 54 years or elder are encouraged to undergo breast screening every 2 years [3]. 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 mammography [4]. Through the use of second opinion or reading by a computer-aided diagnosis (CAD) system, the overall accuracy as well as the false positives and negatives of mass detection, segmentation, and classification could be improved [2]. In the literature, several conventional CAD systems have been developed separately for breast lesion detection, segmentation, or classification [5]. However, there are few studies over a completely integrated CAD system including detection, segmentation, and classification all together. Firstly, detection of breast lesion is an important initial stage to identify the potential region of interest (ROI) of breast lesion in any CAD system. In fact, detection task is still challenging due to the variation of the breast lesions within the surrounding tissues in terms of shape, texture, size, and location in the mammograms [6]. Recently, novel detection approaches based on deep learning were introduced into a CAD system to overcome the challenging tasks of mass detection from mammograms [7]. Secondly, segmentation of breast lesions plays a critical role to accurately extract the specific shape of breast lesions excluding other surrounding normal tissues [8]. Many studies involving mass segmentation have utilized region growing, active contour, and Chan-Vese methods [8]. 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 and different density distribution of the breast lesions. Recently, a few deep learning studies have presented as alternative methodologies for breast lesions segmentations. Indeed, deep learning models have capability to directly extract deep high-level hierarchy feature maps from the input image [8]. Lastly, the majority of CAD systems have been developed to classify the manual extracted breast lesions as either benign or malignant utilizing conventional classifiers [9]. To build such systems, a set of hand-crafted or semi-automatic features describing the characteristics of breast lesions are required. In fact, these conventional CAD systems suffered due to the high degree of similarity of different breast tissues [10]. Alternatively, a few deep learning CAD systems have recently been produced to handle the breast lesion classification task [5]. These systems can learn and extract deep features from input mammograms to achieve better classification performance.

A fully integrated CAD system based on deep learning detection, segmentation, and classification is presented in this chapter. The rest of this chapter is organized as follows. First, You-Only-Look-Once (YOLO) is adopted and

used to detect the breast lesions [11]. Second, a newly deep learning model of full resolution convolutional network (FrCN) is produced for breast lesion segmentation. One advantage of FrCN is to preserve the high resolution of feature maps especially at object edges. Finally, three deep learning models including CNN, ResNet-50, and InceptionResNet-V2 are separately adopted and used to distinguish benign and malignant breast lesions. The presented fully integrated CAD system is evaluated using a public INbreast dataset [12] and its performance is compared with the latest deep learning methodologies of detection, segmentation, and classification.

Related Work

Diagnosis of breast cancer via a CAD system could be improved using the capability of deep learning to accurately represent the high-level deep features of breast lesions in mammograms [4]. Detection of breast lesions is an important task to detect the potential abnormalities for any CAD system [4]. Unfortunately, this task is still a big challenging for researchers and has not been fully resolved yet [8]. Previously, manual detection of breast lesions was widely used in building CAD systems even for deep classification CNN [13]. Most of these CAD systems achieved better classification performance against the traditional machine learning techniques [14]. However, the need to automatically detect breast lesions was recently addressed in the several studies [7, 8, 14]. Few studies based on deep learning present automatic methods to detect breast lesions from the input mammograms [8, 14]. Our preliminary detection results of breast lesions via YOLO utilizing the Digital Database for Screening Mammography (DDSM) are presented in [15]. The detection performance via YOLO was better in comparison to the latest deep learning detection methods [8, 11]. In [16], a new deep learning model called region-based CNN (R-CNN) was used to automatically detect the breast lesions [16]. In [8], another automatic method using a very complex cascade of deep learning models was introduced to detect breast lesions based on R-CNN.

For segmentation, several conventional works have been presented to segment the boundaries of breast lesions from the X-ray mammography images such as growing regions, active contour, and Markov random field (MRF) [17]. However, all of these approaches have shortcomings because they required a prior knowledge of breast lesion contours. Recently, few deep learning studies have been presented and achieved better segmentation results for semantic and medical images [8]. These segmentation models are introduced by adopting and converting the classification functionality of VGG-16 to the segmentation purpose. However, these models suffer from the loss of spatial resolution of the generated feature maps due to the multiple layers of max-pooling and subsampling. Although the utilized max-pooling and subsampling layers reduce the dimensions of derived feature maps and minimize the computation expenses, the spatial resolution of those maps are exponentially decreased [5]. In [18], deep learning model of FCN was used to segment skin lesions from dermoscopy images [18]. Inspired by the structure of FCN, another segmentation model called U-Net was introduced to segment neural brain images including encoder and decoder deep convolutional layers [19]. The extracted feature maps from each encoder layer were combined with the corresponding one in the decoder network. Then, decoder up-sampling and deconvolutional layers were performed to overcome the resolution loss of feature maps [19]. In [20], a deep learning segmentation model called SegNet is presented to segment the semantic images. The SegNet model also consisted of encoder and decoder convolutional network stages. Despite the promising segmentation results of these models, they have not yet been applied for breast lesion segmentation. Up to date, only few attempts have been presented for breast lesion segmentation from mammograms based on deep learning.

Breast lesion classification is the last stage of any CAD systems. The aim of this stage is to recognize or classify the breast lesions as either benign or malignant. Indeed, the performance of classification process mainly depends on the

efficient representations of the derived features [5]. In 2017, Yu et al. presented that a very deep learning model achieved better classification performance with relatively similar computational cost comparing the shallower models [18]. Recently, few integrated CAD systems based on deep learning have been introduced including detection, segmentation, as well as classification for breast lesions [5, 8, 14]. In [21], a hybrid CAD system was introduced based on the combination of deep and hand-crafted features using CNN-based conditional random forest (CRF) to detect the potential lesion ROIs, region growing based on active contour to segment the lesion boundaries, and CNN to classify the breast lesions. This CAD system achieved diagnosis performance in terms of AUC with 94.10%. In [8, 14], an integrated CAD system for breast cancer detection, segmentation, and classification was presented. For breast lesion detection, a complex cascade structure of deep learning was utilized involving multi-scale deep belief network (DBN) with Gaussian mixture (GM) classifier, two stages of R-CNN, two stages of CRF classifier, and a refinement algorithm based on R-CNN [8]. For segmentation, another cascade of deep learning techniques was utilized involving two stages of DBN-based CRF and a refinement method using Chan-Vese active contour [8]. For classification, a regular feedforward version of CNN was used to classify the breast lesions as either benign or malignant. Despite the successes of these CAD systems for breast cancer diagnosis, the remaining challenges still exist including high complexities of memory, practical implementation, and long prediction time.

Materials and Methods

Our integrated CAD system for breast cancer diagnosis includes detection, segmentation, and classification in a single framework. First, an automatic mass detection based on YOLO is performed. Then, a novel deep learning FrCN segments breast lesions. Finally, an automatic breast lesion classification is performed via the convolutional neural networks. A schematic diagram of the integrated CAD system is depicted in Fig. 1.

Dataset

A public X-ray mammography database, INbreast [12], is used in training and evaluation of our integrated CAD system. The classification label and localization ground truth (GT) of the breast lesions for all mammograms in the INbreast database are available and accurately annotated by the experts [12]. All mammograms were collected to represent real breast data with pixel size of 70 μm (microns), and contrast resolution of 14-bit. According to the breast size of the patient, the mammogram size is 3328 × 4084 or 2560 × 3328 pixels. The INbreast dataset includes 410 mammograms (i.e., normal, benign, and malignant) with both views of MLO and CC

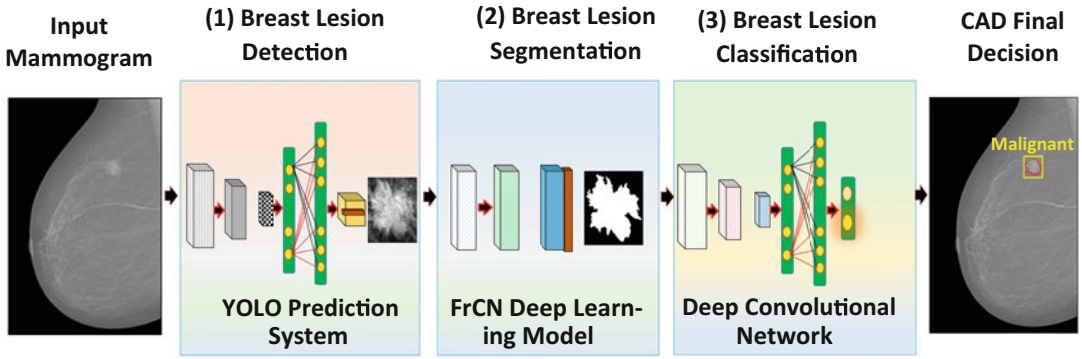


Fig. 1 Schematic diagram of the fully integrated computer-aided diagnosis (CAD) system based on deep learning detection, segmentation, and classification for breast lesions

from 115 patients (cases) [12]. From 90 cases two images (MLO and CC views) are collected for each breast, and from 25 cases only two views (MLO and CC) from one breast having cancer were collected. To evaluate our CAD system, we utilized all mammograms having lesions or masses (benign and malignant) from both views to collect in a total of 107 cases (34 benign and 73 malignant cases) [12]. In INbreast dataset, only two benign and three malignant cases contain two lesions. For benign cases, both CC views are collected from two different patients. For malignant cases, two mammogram views of CC and MLO are collected from the same breast of patient ID: "d713ef5849f98b6c_MG_L," while one CC view is collected from the different patient with ID: "45c7f44839fd9e68_MG_R." Both breast lesions that are visible on both CC and MLO views are used in this study. Thus, 36 benign breast lesions with BI-RAD $\in \{2, 3\}$ and 76 malignant breast lesions with BI-RAD $\in \{4, 5, 6\}$ are collected. Because some of benign and malignant cases have more than one breast lesion, thereby, a total of 112 breast lesions are collected [5, 8, 14].

Datasets Preparation: Training, Validation, and Testing

For data preparation in this work, we randomly divide benign and malignant breast images into three groups: 70% (25 benign and 53 malignant) for training, 10% (4 benign and 8 malignant) for validation, and 20% (7 benign and 15 malignant) for testing as previously performed [8, 14]. In addition, unbiased double cross-validation strategy is utilized as follows. Trainable parameters of the proposed deep learning models are optimized during the training process using only training and validation datasets [14]. Then, the final performance of the presented CAD system is only evaluated using the testing dataset. In fact, double cross-validation is very important for parameters optimization and selections due to the following reasons. First, to be sure testing dataset is totally isolated during the training process. Second, to avoid any bias that may occur during the training process. Third, to ensure that the overall performance of the presented CAD system is robust and reliable for real testing as well. In this study, fivefold cross-validation tests are also carried out with training, validation, and testing sets, which are generated by stratified partitioning to ensure that each breast image gets tested equally and to prevent any bias error for training and testing tasks. For each fold, those breast lesions that are visible on the same mammogram view or on both CC and MLO views from the same patient should be categorized in one set of training, testing, or validation to avoid the system bias as well.

Preprocessing

Preprocessing of all mammograms is achieved using the following steps. First, Otsu's thresholding is used separating the breast region from its background to exclude the unwanted information [5]. Second, contrast limited adaptive histogram equalization (CLAHE) technique has been successfully used to enhance the image contrast between the suspicious lesions and their surrounding normal tissues [5, 8]. Indeed, CLAHE is an image contrast enhancement method which depends on the histogram equalization process [5]. There are two sequential steps to apply CLAHE for breast image enhancement. First, the histogram of the entire mammogram is divided into multi-regions at certain thresholds. Then, the histogram equalization process is locally applied over each region. Same preprocessing strategy is used for all mammograms that used in this study without subsampling.

Data Balancing and Augmentation

To develop deep learning models, a large amount of annotated dataset is required for parameter optimization and selection during training process [7, 8, 14]. To avoid any bias during training process, data balancing and augmentation are widely used as regularization strategies for deep learning models. Indeed, data balancing and augmentation strategy are applied only for training dataset [5, 8, 14]. That means the original breast lesion and its many representations (i.e.,

augmented data) are only included in the training set. Whereas, the testing set is only used without augmentation to evaluate the proposed CAD system over detection, segmentation, and classification stages. As aforementioned, the training dataset from INbreast is unbalanced containing 25 benign and 53 malignant cases. Balancing of training dataset means generating almost an equal number of breast images from both benign and malignant classes. Due to that, all breast benign mammograms from the INbreast training dataset are vertically flipped to balance benign and malignant cases. Thus, 103 mammograms (i.e., 50 benign and 53 malignant) are generated. Then, the balanced training datasets are augmented 22 times using the following strategies. First, all mammograms are rotated eight times around the origin center with the angle of $\Delta\theta = 45°$ (i.e., 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°) [7, 8]. Second, left-right and up-down flipping are applied for all rotated mammograms with 90° and 270°. Third, random scaling and translations are applied ten times for all original breast images. In the total, 2266 mammograms are generated from INbreast to train our CAD system over five-fold cross-validation for all stages of detection, segmentation, and classification. For each k-fold, the strategies of data splitting, balancing, and augmentation are applied in the same ways.

Initialization of Trainable Parameters for Deep Learning Models

To accelerate and avoid overfitting that may occur during training process of all deep learning models, the trainable parameters of convolutional and fully connected (FCs) layers should be initialized [5, 8]. In the literature, several strategies to initialize the trainable parameters of deep learning models were used such as random initialization and transfer learning [5, 8, 14]. In this study, transfer learning is used to initialize the parameters in two consecutive steps. First, all deep learning models are pre-trained using a large annotated computer vision dataset (i.e., ImageNet [22]). Second, these models are fine-tuned using our augmented annotated dataset (i.e., mammograms).

Breast Lesion Detection via YOLO

Detection of breast lesion is the first critical task for the CAD system to identify all potential lesions from entire mammograms. In this chapter, we adopt and use a deep learning model called YOLO, a regional ROI-based CNN technique, to perform the detection task [11]. Our preliminary works using YOLO has proven that this technique is effective for breast lesion detection tasks using a public X-ray mammography dataset, DDSM [5, 15]. Since the INbreast dataset includes accurate ground truth, YOLO could be a good choice for detection of breast lesions due to the following reasons. First, YOLO has a robust ability to directly detect the breast lesions from the entire mammograms [15]. Second, detected bounding boxes via YOLO accurately align the breast lesions, thereby, a low rate of false positives is achieved compared with other studies [14]. Third, YOLO can detect the most challenging cases of breast lesions even when they exist over the pectoral muscles or inside dense regions. Fourth, required testing time and memory are extremely lower than other more complex deep learning models [14].

Breast Lesion Segmentation via FrCN

Once the breast lesions are detected from the previous detection stage of our CAD system, they are directly passed into the novel segmentation model, FrCN, to segment the breast lesions end-to-end. FrCN composes of two main consecutive encoder and decoder networks. The encoder network involves thirteen convolutional layers, while decoder is built using three convolutional layers. Unlike the previous deep learning models, the max-pooling and subsampling layers are removed from the both encoder and decoder networks to preserve the full spatial resolution of the original input image (i.e., breast lesion) as well as the details of the objects. This is a key modification to prevent any information loss during feature map generation. Therefore, the high-level deep feature maps in each block are generated utilizing only the convolutional process, preserving the full

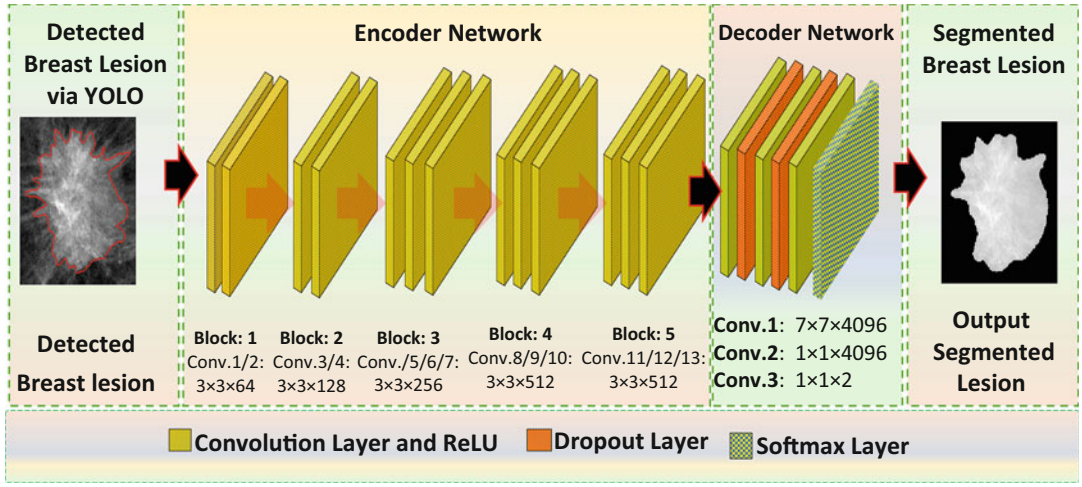


Fig. 2 Segmentation deep learning model of a full resolution convolutional network (FrCN). The input image is a detected breast lesion ROI highlighted with its ground truth (red), while the output image is the corresponding segmented ROI

resolution of the input images. By this modification, FrCN is able to maintain the details and edges especially for the tiny objects. Because the convolutional layers on the full resolution of the input images without subsampling in the encoder network are used, up-sampling and deconvolutional layers in the decoder network are not required. The final output of deep feature maps is directly passed into a softmax function to obtain the probability for each image pixel. Finally, a non-linear activation function of rectified linear unit (ReLU) is utilized after each block in the encoder and decoder networks. The schematic diagram of our deep learning FrCN segmentation model is shown in Fig. 2. To evaluate the overall segmentation performance of FrCN, a direct comparison against other existing deep learning models such as FCN [23], SegNet [20], and U-Net [19] is presented using the same dataset from the INbreast database [12].

Breast Lesion Classification via Three Convolutional Neural Networks

Once breast lesions are detected as well as segmented, deep learning models of our feedforward CNN [5], ResNet-50 [24], and InceptionResNet-V2 [25] are separately used to classify the breast lesions as benign or malignant. These deep learning models are used to perform the classification task of our CAD system. Indeed, we use the regular feedforward CNN presented in our recent study [5]. Also the deep learning models of ResNet-50 and InceptionResNet-V2 are adopted replacing their last two layers by other four layers of global average pooling (GAP) layer, two dense layers with ReLU activation functions, and a logistic regression layer of softmax. Deep learning models such as ResNet-50 and InceptionResNet-V2 are recently introduced with very deep convolutional layers to improve the classification performance preserving the computational burden to be similar as in the shallower CNNs [26]. The main principle of the Inception networks is to produce multiple feature maps from the input images using different parallel pathways with different convolutional filters [26]. By using these remedies with the Inception models, their execution time overcomes other state-of-the-art deep learning models of CNN and ResNet-50 [24]. Since the residual connections are inherently important to train very deep architectures, the filter concatenation stage of Inception models is replaced by the residual ones producing the InceptionResNet models [26]. Training of the Inception models is significantly accelerated using the residual connections as demonstrated in [26].

Meanwhile, the classification performance of the residual Inception models is slightly better comparing the Inception modules without the residual connections [26].

Experimental Settings

For evaluation of the CAD system, fivefold cross-validation tests are carried out in each stage of the CAD system using training, validation, and testing datasets. These sets are generated by stratified partitioning to ensure that each mammogram gets tested equally preventing any bias error [14]. To avoid any bias that may occur during training process, weights optimization process utilizing a weighted cross-entropy loss function as well as double cross-validation are used [5, 14].

Detection Experimental Settings

Due to the different sizes of both mammogram views, all breast images are resized into a fixed size of 448 × 448 pixels [5]. The YOLO-based CAD system is trained for 135 epochs and mini-batch size of 64 using only training and validation datasets. Also, stochastic gradient descent (SGD) optimizer is used with momentum and decay of 0.9 and 0.0005, respectively. The breast lesions are considered to be correctly detected if the intersection over union ($IoU_{GT}^{Ext.}$) between the extracted and ground truth bounding boxes is greater than or equals 50%. Moreover, the false positive candidates of breast lesions are manually excluded before the segmentation and classification stages of the CAD systems as previously applied in [5, 8, 14]. This is because there is a lack of ground truth information of falsely detected lesions to derive the performance evaluation metrics especially for the segmentation stage [8]. Thus, the evaluation results of segmentation and classification tasks are computed with the exception of the falsely detected cases of breast lesions.

Segmentation Experimental Settings

Similar to the detection stage, the same fivefold cross-validation is performed for all segmentation deep learning models: FCN [23], SegNet [20], U-Net [19], and FrCN. To train all of these deep learning models, Adam optimizer is used with a learning rate of 0.001. Meanwhile, 135 epochs and 20 mini-batches are used to optimize and select the model parameters with the training and validation datasets. 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 overfitting [5].

Classification Experimental Settings

All detected and segmented breast lesions are normalized and resized using bi-cubic interpolation into a fixed size of 128 × 128, 224 × 224, and 299 × 299 pixels for CNN, ResNet-50, and InceptionResNet-V2, respectively. Then, all these breast lesions are directly fed into the classification stage producing the final prediction of our CAD system. In fact, these deep learning models are adopted to compare the recognition performance of shallower CNN against the deeper models of ResNet-50 and InceptionResNet-V2. This comparison is performed under the same training and testing settings for all deep learning models. To verify the CAD system for classification, the same fivefold cross-validation is performed similar to the detection and segmentation stages. For training, Adam optimizer with the initial learning rate of 0.0001 and weight decay of 0.0005 is used. The learning rate is reduced by 50% if the loss function does not decrease by 0.001 every 10 epochs. The mini-batch size and number of epochs are set to 24 and 130, respectively. Dropout of 0.3 is used for both fully connected layers in all deep learning models to prevent overfitting as well as accelerate the training process [5, 8].

Implementation Environment

All these experiments are performed on a PC with the following hardware specifications: Intel(R) Core (TM) i7-6850K with 16 GB RAM, clock speed 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 [27] and Keras [28] deep learning libraries, while the detection and classification models are implemented under the Tensorflow environment [29].

Experimental Results and Discussion

Evaluation Metrics

Each detection, segmentation, and classification stage of the CAD system gets separately evaluated using an overall accuracy (Acc.), sensitivity (Sen.), specificity (Sep.), F1-score (Dice), Jaccard (Jac.), and Matthews correlation coefficient (MCC). Moreover, area under ROC curve (AUC) is also used to evaluate the CAD system over segmentation and classification stages. The definition and criteria for all of these evaluation metrics are available in [5].

Breast Lesion Detection Results

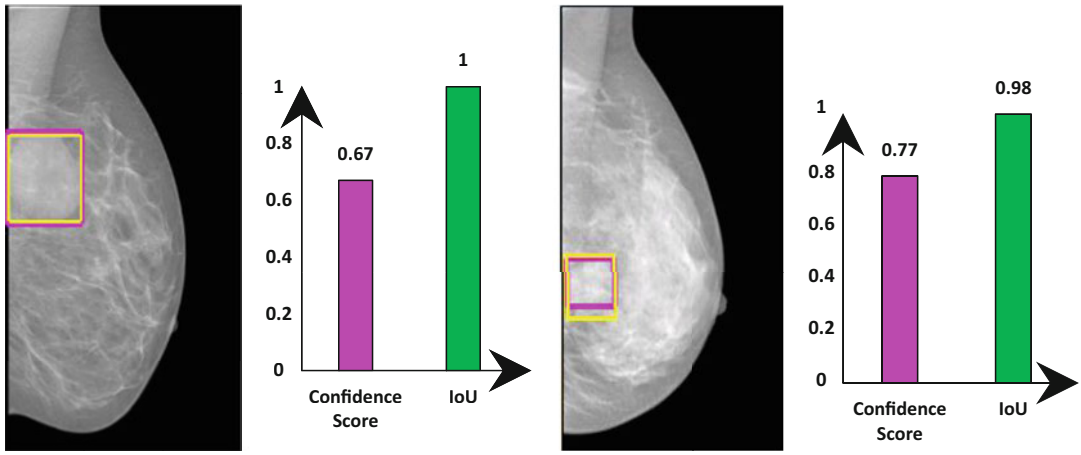
The detection performance of breast lesions via YOLO over fivefold cross-validation tests with the testing dataset is reported in Table 1. The false detection cases presented in Table 1 indicate the cases when $\mathrm{IoU}_{\mathrm{GT}}^{\mathrm{Ext.}} < 50\%$. This means that the final extracted breast lesion has no enough overlap ratio with its GT [5, 7, 8]. Fortunately, YOLO detects at least one bounding box indicating breast lesion from all testing mammograms. In this study, the false detection cases are excluded over each test fold for the next stages of segmentation and classification. An average overall detection accuracy of 97.27% at 0.25 false positive per image (FPI), MCC of 93.93%, and F1-score of 98.02% present the reliable detection performance of the YOLO detector. Examples of the qualitative breast lesion detection results via YOLO identifying the potential breast lesion ROIs are shown in Fig. 3. It is clearly shown that YOLO can accurately detect and align the detected bounding boxes surrounding the breast lesions with high prediction confidence score and high overlapping ratio of IoU. In fact, confidence score indicates the probability of the presence of breast lesions, while overlapping ratio indicates how much the lesion detection localization is accurate. Moreover, a comparison of the detection results using YOLO against the other latest deep learning methods is listed in Table 2. It is clearly shown that YOLO achieved much better detection accuracy with higher prediction speed in comparison to other deep learning detection methods. Also, YOLO can detect even the most challenging cases when the breast lesions exist over the pectoral muscles or inside the breast dense tissues as depicted in Fig. 3(a) and (b), respectively. Therefore, the YOLO detector plays a critical role in the CAD system, achieving the best detection performance of breast lesions comparing the latest deep learning models.

Breast Lesion Segmentation Results

The average segmentation performance results of our FrCN segmentation model against FCN, SegNet, and U-Net over fivefold tests are reported in Table 3. For comparison, the results of all deep learning models are achieved without any refining pre- and/or post-processing. The quantitative measurements of all metrics are computed per pixel of the segmented maps with the same resolution of the input detected breast lesions. FrCN obviously outperformed other methods with an average Dice index of 92.36%, Jaccard coefficient of 85.81%, overall accuracy of 92.69%, and MCC of 85.36%. U-Net achieved better segmentation results comparing SegNet in terms of all evaluation metrics. In addition, SegNet achieved better segmentation performance in terms of specificity with 96.38%.

Table 1 The detection performance of the breast lesions over fivefold cross-validation using YOLO detection model on the test sets from the INbreast dataset

	Benign		Malignant		Total		Metrics (%)		
Fold test	True	False	True	False	True	False	Acc.	MCC	F1-score
1st Fold	6 85.71%	1 14.29%	15 100%	0 0.0%	21 95.45%	1 4.55%	95.45	89.64	96.77
2nd Fold	7 100%	0 0.0%	14 93.33%	1 6.67%	21 95.45%	1 4.55%	95.45	90.37	96.56
3rd Fold	7 100%	0 0.0%	15 100%	0 0.0%	22 100%	0 0.0%	100	100	100
4th Fold	6 85.71%	1 14.29%	15 100%	0 0.0%	21 95.45%	1 4.55%	95.45	89.64	96.77
5th Fold	7 100%	0 0.0%	15 100%	0 0.0%	22 100%	0 0.0%	100	100	100
Avg.(%)	94.28	5.71	98.67	1.33	97.27	2.73	97.27	93.93	98.02



(a) (b)

Fig. 3 Examples of detected breast lesions from the INbreast dataset via the YOLO detector. Detected breast lesions exist over the pectoral muscle and inside the breast dense tissues as shown in (**a**) and (**b**), respectively. Detected breast lesions (magenta) and their ground truths (yellow) are superimposed on the original mammograms

Table 2 The detection performance comparison between the YOLO detector against the other latest studies on the test sets from the INbreast

Reference	Method	Dataset	Prediction time per image (Sec.)	Detection accuracy (%)
Dhungel et al. [8], Carneiro et al. [14]	Cascade deep learning F-RCN, DBN, and CRF	INbreast	39	90.0 at 1.0 FPI
Kozegar et al. [30]	Adaptive threshold with some of machine learning techs	INbreast	108	87.0 at 3.67 FPI
Our presented method	YOLO-based CAD system	INbreast	0.014	97.27 at 0.25 FPI

Table 3 The average breast lesion segmentation performance over fivefold cross-validation on the test sets from the INbreast dataset

		Evaluation metrics (%)						
Fold test	Deep learning method	Dice	Jac.	Sen.	Spe.	Acc.	AUC	MCC
Avg. (%)	FCN	88.05	79.14	82.06	95.10	88.95	89.52	80.30
	SegNet	89.40	81.83	83.52	96.83	90.26	90.26	82.04
	U-Net	90.98	83.77	87.03	95.70	91.91	91.72	83.24
	Our FrCN	92.36	85.81	92.94	92.47	92.69	92.70	85.36

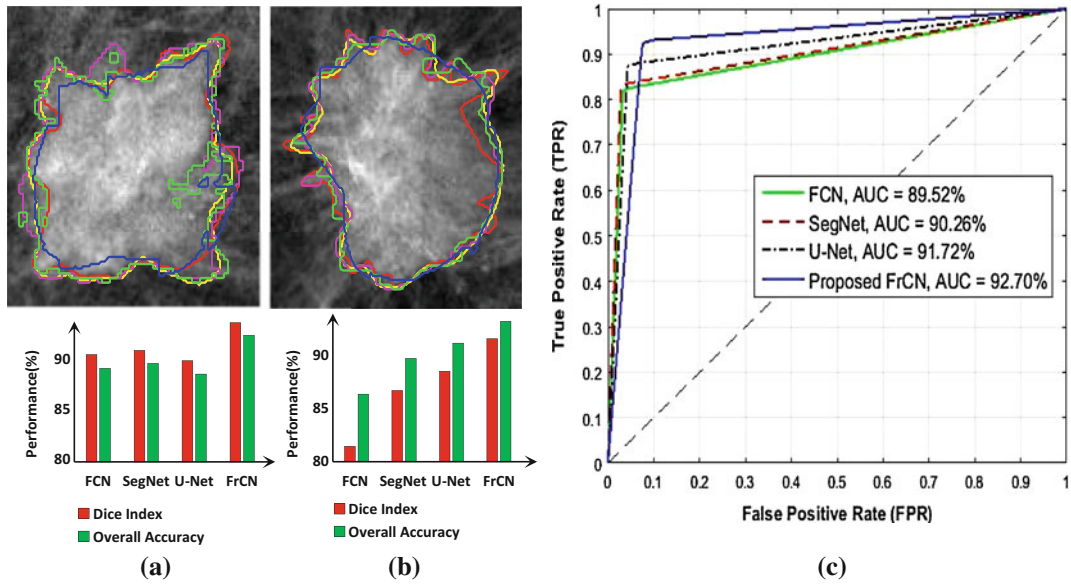


Fig. 4 Examples of qualitatively breast lesion segmentation results via our deep learning model of FrCN against the FCN, U-Net, and SegNet are shown in (**a**) and (**b**), while (**c**) presents the evaluation performance of all deep learning models in terms of ROC curves with their AUCs. The counters in (**a**) and (**b**) indicate the ground truth (red), FrCN (yellow), U-Net (green), SegNet (magenta), and FCN (blue)

Moreover, examples of the qualitative breast lesion segmentation results for FrCN against FCN, SegNet, and U-Net are depicted in Fig. 4 (a) and (b). Moreover, the segmentation performance of FrCN against all other methods is evaluated by the AUC over all test folds. Figure 4(c) shows an example of the ROC curves with AUCs from the 2nd test fold for comparison among all segmentation models. As presented in Table 3 and Fig. 4(c), the performance of breast lesion segmentation via FrCN outperformed all other methods in terms of AUC with 92.70%. In addition, FrCN achieved faster training time with 6.42 h than FCN, SegNet, and U-Net with 12.94, 6.81, and 8.03 h, respectively. For testing, FrCN segments the individual breast lesion in 8.51 s comparing 10.25, 10.48, and 10.66 s for FCN, SegNet, and U-net, respectively. Despite U-Net achieved better segmentation results than SegNet, but it is slightly slower to perform training and testing tasks. FrCN overcomes the limitations of the latest deep learning segmentation models in terms of preserving high resolution and better performance for large and tiny objects.

Breast Lesion Classification Results

Once the detected lesions are segmented via FrCN, deep learning convolutional networks of CNN, ResNet-50, and InceptionResNet-

Table 4 Breast lesion classification performance as an average over fivefold cross-validation on the test sets from the INbreast dataset

		Benign class			Malignant class			Weighted metrics		
Fold test	Method	Sen.	Spe.	Dice	Sen.	Spe.	Dice	ACC	AUC	MCC
Avg. (%)	CNN	84.75	90.57	82.19	90.57	84.76	91.77	88.74	87.67	74.17
	ResNet-50	91.42	93.24	88.35	93.24	91.43	94.52	92.56	92.33	84.10
	Inception ResNet-V2	90.47	97.33	92.16	97.33	90.47	96.64	95.32	93.91	89.39

Table 5 Comparison performance classification results of the fully integrated CAD system against others through all stages of detection, segmentation, and classification

Reference	Prediction classes	Prediction time per image (Sec.)	Overall classification accuracy and (AUC) (%)	Hardware specs
Dhungel et al. [8]	Benign/Malignant	41	91 and (76)	Intel Core i5-2500k, 8GB RAM, 3.30 GHz, and GPU of NVIDIA GeForce GTX 460 SE 4045 MB
Carneiro et al. [14]	Normal/Benign/ Malignant	41	NA and (78) → Benign Vs. Malignant NA and (86) → Malignant Vs. (Normal + Benign)	Intel Core i7, 8GB RAM, 2.3 GHz, and GPU of NVIDIA GeForce GT 650M 1024 MB
Our CAD system	Benign/Malignant	9.32	95.32 and (93.91)	Intel Core i7-6850K, 16GB RAM, 3.360 GHz, and GPU of NVIDIA GeForce GTX 1080

V2 are used to classify these lesions as either benign or malignant. Table 4 shows the average breast lesion classification performance for each class of benign and malignant over fivefold tests. The evaluation results presented in Table 4 are computed for the correctly detected breast lesions. It is obviously noted that the deep learning model of InceptionResNet-V2 achieved better classification results in terms of sensitivity of 97.33%, specificity of 90.47%, overall accuracy of 95.32%, F1-score of 94.40%, and AUC of 93.91%, whereas the shallower model of CNN achieved the lowest classification performance results in comparison to other deeper models of ResNet-50 and InceptionResNet-V2 in terms of all evaluation metrics. This means that the deeper models achieve better classification performance against the shallower ones. However, the promising classification performance of our fully integrated CAD system is achieved due to many reasons as follows. First, the potential breast lesion ROIs are accurately detected and aligned using the prediction model of YOLO. Second, the robust segmentation deep learning model of FrCN plays a critical role to extract the specific region of breast lesions minimizing the false positive and negative pixels from the surrounding normal tissues. Third, the high deep level feature maps derived using the state-of-the-art deep learning models highly contribute to improvement of the overall diagnostic performance of the CAD system. Finally, a comparison between our fully integrated CAD system with respect to the latest studies based on the deep learning is presented in Table 5. All these studies are evaluated using the INbreast dataset through each stage of detection, segmentation, and classification. It is clearly shown that our CAD-based deep learning could handle all these stages achieving a higher performance as well as much faster prediction time. Therefore, the promising prediction performance of the CAD system seems to make it more feasible towards practical applications.

Conclusion

In this chapter, a fully integrated CAD system based on deep learning including detection, segmentation, and classification is presented for au-

tomatic diagnosis of the breast lesions in a single framework. To automatically detect breast lesions from the entire mammograms, the YOLO-based lesion detection could be used. Due to the segmentation capability of the FrCN model, the CAD system could achieve a much better diagnostic results. Hence, the detection and segmentation end-to-end of breast lesions could be a key to minimize the false positive and negative rates and then improve the overall performance of our integrated CAD system. Moreover, classification based on the convolutional deep learning contributes to accurately classify breast lesions. A fully integrated CAD system based on deep learning methodologies could be beneficial for practical applications of future medical imaging systems.

Acknowledgements This work was supported by International Collaborative Research and Development Programme funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) (N0002252). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (2019R1A2C1003713).

References

1. Siegel RL, Miller KD, Jemal A (2017) Cancer statistics, 2017. CA Cancer J Clin 67(1):7–30
2. Al-antari MA, Al-masni MA, Park SU, Park JH, Kadah YM, Han SM, Kim T-S (2016) Automatic computer-aided diagnosis of breast cancer in digital mammograms via deep belief network. In: Global conference on engineering and applied science (GCEAS), Japan, pp 1306–1314
3. Jill J (2015) Breast cancer screening guidelines in the United States. JAMA 314:1658–1658
4. Al-antari MA, Al-masni MA, Park SU, Park JH, Metwally MK, Kadah YM, Han SM, Kim T-S (2017) An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network. J Med Biol Eng. https://doi.org/10.1007/s40846-017-0321-6
5. Al-antaria MA, Al-masni MA, Choi M-T, Han S-M (2018) A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. Int J Med Inform 117:44–54
6. Casellas-Grau A, Vives J, Font A, Ochoa C (2016) Positive psychological functioning in breast cancer: an integrative review. Breast 27:136–168
7. Al-Masni MA, Al-Antari MA, Park JM, Gi G, Kim TY, Rivera P, Valarezo E, Choi MT, Han SM, Kim TS (2018) Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Comput Methods Programs Biomed 157:85–94
8. Dhungel N, Carneiro G, Bradley AP (2017) A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med Image Anal 37(1):114–128
9. Yassin NI, Omran S, Houby EM, Allam H (2018) Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. Comput Methods Programs Biomed 156:25–45
10. Chakraborty J, Midya A, Rabidas R (2018) Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns. Exp Syst Appl 99(1):168–179
11. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: 2016 IEEE conference on computer vision and pattern recognition, pp 779–788
12. Moreira I, Amaral I, Domingues I, Cardoso A, Cardoso M, Cardoso J (2012) INbreast: toward a full-field digital mammographic database. Acad Radiol 19(2):236–248
13. Jiao Z, Gao X, Wang Y, Li J (2016) A deep feature based framework for breast masses classification. Neurocomputing 197(C):221–231
14. Carneiro G, Nascimento J, Bradley AP (2017) Automated analysis of unregistered multi-view mammograms with deep learning. IEEE Trans Med Imaging 36(11):2355–2365
15. Al-masni MA, Al-antari MA, Park JM, Gi G, Kim TY, Rivera P, Valarezo E, Han S-M, Kim T-S (2017) Detection and classification of the breast abnormalities in digital mammograms via regional Convolutional Neural Network. In: 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC'17), Jeju Island, South Korea, 2017, pp 1230–1236
16. Ayelet A-B, Karlinsky L, Alpert S, Hasoul S, Ben-Ari R, Barkan E (2016) A region based convolutional network for tumor detection and classification in breast mammography. In: International workshop on large-scale annotation of biomedical data and expert label synthesis. Springer, Athens, pp 197–205
17. Cardoso JS, Domingues I, Oliveira HP (2015) Closed shortest path in the original coordinates with an application to breast cancer. Int J Pattern Recognit Artif Intell 29(1):2
18. Yuan Y, Chao M, Lo Y-C (2017) Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance. IEEE Trans Med Imaging 36(9):1876–1886
19. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention

20. Badrinarayanan V, Kendall A, Cipoll R (2016) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561
21. Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, den Heeten A, Karssemeijer N (2017) Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 35:303–312
22. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: 25th international conference on neural information processing systems, USA, 2012, pp 1097–1105
23. Shelhamer E, Long J, Darrell T (2017) Fully Convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39(4):640–651
24. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. arXiv:1602.07261v2, pp 770–787
25. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprintarXiv:1409.1556
26. Szegedy V, Ioffe S, Vanhoucke V (2016) Inception-v4, Inception-ResNet and the impact of residual connections on learning. arXiv:1602.07261v2 [cs.CV]
27. Lab L (2017) Theano. University of Montreal [Online]. Available: http://deeplearning.net/software/theano/tutorial/. Accessed 10 2017
28. Chollet F (2017) Keras: the Python deep learning library. MIT, [Online]. Available: https://keras.io/. Accessed 10 2017
29. Google Brain Team, TensorFlow, 9 11 2017. [Online]. Available: www.tensorflow.org. Accessed 10 2017
30. Kozegar E, Soryani M, Minaei B, Inês D (2013) Assessment of a novel mass detection algorithm in mammograms. J Cancer Res Ther:592–600