Utilization of DenseNet201 for diagnosis of breast abnormality

Xiang Yu[1], Nianyin Zeng[3,*], Shuai Liu[4,*], Yu-Dong Zhang[1,2,*]
Co-Corresponding Emails: zny@xmu.edu.cn, cs.liu.shuai@gmail.com, yudongzhang@ieee.org

1 Department of Informatics, University of Leicester, Leicester, LE1 7RH, UK
2 School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, P.R. China
3 Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen, 361005, P.R. China
4 College of Information Science and Engineering, Hunan Normal University, Changsha, 410000, P.R. China

Abstract

As one of the leading killers of females, breast cancer has become one of the heated research topics in the community of clinical medical science and computer science. In the clinic, mammography is a publicly accepted technique to detect early abnormalities such as masses and distortions in breast leading to cancer. Interpreting the images, however, is time-consuming and error-prone for radiologists considering artificial factors including potential fatigue. To improve radiologists' working efficiency, we developed a semi-automatic computer-aided diagnosis system to classify mammograms into normality and abnormality and thus to ease the process of making a diagnosis of breast cancer. Through transferring deep convolutional neural network DenseNet201 on the basis of suspicious regions provided by radiologists into our system, we obtained the network we termed as DenseNet201-C, which achieved a high diagnostic accuracy of 92.73%. The comparison results between our method and the other five methods show that our method achieved highest accuracy.

Keywords: Breast cancer; Transfer learning; DenseNet201-C

1 Introduction

Breast cancer remains one of the common cancers worldwide while it has taken millions of lives away. Risk factors inducing cancer include family history and personal factors such as lack of physical exercise, obesity and so on [1]. While using certain medicines and controlling methods do decrease the risk of breast cancer at the prevention stage, early detection and diagnosis can be an effective way to reduce the mortality rate and thus control cancer. To achieve early detection, experts in the field used mammography technique, which turned out to be a mainstream technique in this field. With the help of mammography, radiologists are able to visualize the abnormalities in the breast. However, due to the overlapping tissues and high complexity of the image, sometimes double reading is required given that the first interpreting procedure is already time-consuming.

The most significant early syndromes of breast cancer are micro-calcification and mass. Calcifications are mineral deposits in the form of white small spots or clusters in the presence of mammogram images. The clinical practice of diagnosing the calcifications is a challenging task due to the large variations in size, shape, and distribution in mammogram images. While radiologists usually have to spot the calcifications in mammogram images manually, some of the calcifications may be missed due to the distraction of radiologists or subtleness of the calcifications. Mass, or lump, is also another typical form of abnormality, which can be categorized into being benign and malignancy according to the severity. Unlike the spiculated malignant masses, benign masses are usually circumscribed and non-cancerous. But early identification of both types ahead of any further deterioration is still of great significance. Given the challenges in accurate diagnosis, the misdiagnosis is nothing but common and falls into two categories, namely false positives, and false negatives. Inevitably, the false negatives will increase the mortality rate of breast cancer without noti-

fiable control while false positives lead to unnecessary measurements and squandering of resources.

To help radiologists improve their working efficiency as well as the accuracy of diagnosis, considerable previous works on designing CAD systems of breast cancer have been done [12] [20] [24] [31]. There are two main categories of CAD systems according to the functionality: detection and diagnosis. Usually, locating the abnormal areas in images is the basic yet most important step in almost all of the CAD systems. They can be very helpful when it is difficult for radiologists to detect or locate abnormal areas given complicated contents of images. Additionally, diagnosis systems are generally applied afterward to help radiologists make decisions on the malignancy of abnormalities when necessary. To detect the masses in mammograms, Neeraj et al. designed an automated detection system with a cascade of deep learning and random forest classifiers introduced [8]. Deep belief network was used as a classifier to select the multi-scale specious regions at the first stage followed by random forests classifier to determine the regions surviving the first stage to be mass or not. Experiments on two different data sets showed promising results. A similar work to detect calcification was proposed with cascaded classifiers introduced [26]. Also, there is a large number of works on diagnosis systems regarding micro-calcification and mass [5] [4] [24]. A trained nine-layer CNN in the breast cancer detection system, designed by Pan, outperformed six other the state-of-the-art approaches [36]. The diagnosis accuracy reached 94% but is still improvable. In another breast cancer detection system, a novel convolutional neural network (CNN) named GlimpseNet was introduced [13]. Multi regions of interest (ROI) are extracted simultaneously and are then classified. A diagnostic result for the full image is then given by pooling them to give a disgnostic result for the full image. However, as the authors stated, the overfitting problem hasn't been fixed while the performance of accuracy gained 4.1% compared to existing algorithms.

Despite the fact that deep CNNs have been widely used in fields including not only medical analysis but also industrial application [15] [16] [17] [18] [19], some problems remain to be solved in practical usage. Constraints of practical problems such as the limited size of training data refrain the performance of deep CNNs trained from scratch to be satisfactory. Also, as pointed out by Jason [34], the bottom layers in deep CNNs especially the first layer learn some general features similar to Gabor filters or color blobs, but the top layers learn more specific features regarding the data set. Consequently, transfer learning, which is introduced to adopt classifiers trained for other categories to classify certain categories, turns out to be an effective way to solve the dilemmas [6] [9] [29] [32]. There are grossly two ways of transferring existing networks into problem-oriented ones. However, the first procedure of two different transfer learning methods is the same that is to copy the first n bottom layers from base networks to target networks while parameters in the remaining layers of target networks are randomly initialized. The difference between the two methods is that one leaves the parameters in copied layers frozen while the other fine-tunes the parameters when training the target networks. Freezing technique is suggested when the size of the target data set is small while the numbers of parameters are large otherwise fine-tuning should be adopted when large data set accompanied with small numbers of parameters. As such, there are many outstanding jobs of adopting transfer learning to different areas including detection and diagnosis of breast cancer. Ravi et al. transferred a nine-layer deep CNN that was pre-trained on a large mammogram data set but improved the accuracy of detecting mass in digital breast tomosynthesis (DBT) significantly from 0.80 to 0.91 [28]. In another attempt, Benjamin et al. improved the accuracy of mammographic tumor classification from 0.81 to 0.86 after applying transfer learning [22].

In order to advance the diagnosis system of breast cancer, we proposed a CAD diagnosis system to classify abnormalities in mammogram images. For simplicity, we classify regions acquired from mammograms into normality class when no abnormality found while regions containing abnormalities like micro-calcifications and masses are classified into abnormality class. The input of the system is regions cropped from mammogram images according to hand-crafted labels by experts for MINI-MIAS data set [3], which contains 322 mammograms in total. We transferred the state-of-the-art networks DenseNet201 into our system as classifiers [21], which removes rudimentary procedures such as segmentation in traditional diagnosis systems. To specify whether base networks should be frozen or not in this problem, we used two transfer learning methods and compared the results. Also, we validated our system by comparing the accuracy of the system with the state-of-the-art works' and came to the conclusion that our method worked best amongst all of them.

2 DenseNet

Among all of the newly proposed deep CNNs, DenseNet was known for the startling performance on the competitive object recognition benchmark tasks such as ImageNet and CIFAR-100 [7] [25]. However, the surge of deep CNNs can be dated back to 2012. In 2012, a network with novel architecture named AlexNet, the first attempt to solve large scale image classification challenge by deep CNN [2], won the first place and second place in localization and classification task respectively. The depth of AlextNet, only comprising of five convolutional layers and three fully connected layers, may be shallow compared to that of the state-of-the-art deep CNNs' designed in recent years but it indicated a promising future for deep CNNs as can be proved by the explosive growth in the number of new deep CNN models.

For newly developed networks, various shortcut connections turn out to be effective methods to overcome possible problems such as vanishing gradient when training deep CNNs. In [30], an improved inception module composing of short-cut branch and a few deeper branches solved the problem of vanishing/exploding gradients when the networks are going deeper. To ease the training process of deep networks based on gradient, networks termed highway networks were designed to allow information produced by preceding layers flow to subsequent layers without loss of information. Similarly, residual networks proposed by He et al. achieved the same goal but used residual learning method [14]. While taking advantage of deep and wide architectures can be useful techniques to improve the performance of deep CNNs, DenseNet realizes the goal of easy training and parameter efficiency through feature reuse, which inputs the concatenated feature-maps produced by all preceding layers into the subsequent layer. In that way, deep layers in the networks are allowed to access all of the feature-maps produced by previous layers and thus reuse features. That means the feature-maps X^l in the lth layer can be projected by all feature-maps, $X^0 \cdots X^{l-1}$, in previous layers in the form of:

$$X^l = H_l([X^0, \cdots, X^{l-1}]) \tag{1}$$

Where $[X^0, \cdots, X^{l-1}]$ are feature-maps in 0th, $\cdots$, $l-1$th layer respectively. $H_l(\cdot)$ is defined as a composite function of three operations comprising of batch normalization (BN)[23] [10], a rectified linear unit (ReLU) and a 3 × 3 convolution (Conv). In traditional deep CNNs, convolutional layers were generally followed by down-sampling layers that reduce the width and height of feature-maps to half of them. Consequently, concatenation of feature-maps before down-sampling layers and after would be problematic due to different sizes. To solve this problem, dense blocks were craftily designed followed by down-sampling layers while layers in dense blocks are densely connected. As a result, the sizes of feature-maps in dense blocks remained unchanged while were halved after down-sampling. Therefore, for a dense block with L layers, the total number of direct connections between layers is $L(L+1)/2$ whereas the number of connections in a traditional convolutional network with L convolution layers is only L. However, the deeper the layers the larger the number of concatenated feature-maps input to the following layers. If no constraints on the linear growth in the number of feature-maps, huge computation expenditure would be a disaster. Hence growth rate k is designed to control the number of newly produced feature-maps in each layer. As a result, the total number of feature-maps in the lth layer of a dense block is $k_0 + (l-1) * k$, where k_0 is the number of the channels in the input layer. To reduce the complexity of computation, bottleneck layers were introduced by utilizing a 1 × 1 convolution before each 3 × 3 convolution. Besides bottle layers, layers named transition layers serve to control the number of output feature-maps in a certain depth of the networks and thus improve the compactness of whole networks. Fig. 1 shows the process of concatenating and producing new feature-maps in the first dense block of DenseNet201. As transition layer generally appears after dense block in certain depth, there is no transition layer after the first dense block.

3 Methodology

3.1 Data set

In this paper, the data set was MINI-MIAS, a public mammogram database which can be acquired at http://peipa.essex.ac.uk/info/mias.html. This database contains 114 abnormal tissues images and 208 normal images and thus result in 322 images in total with labels of possible lesion regions labeled by radiologists. For images with the abnormalities, they can be categorized into six classes including calcification, well-defined mass, spiculate mass, ill-defined mass, architectural distortion, and asymmetry. Some typical examples of

these classes are presented in Fig. 2.
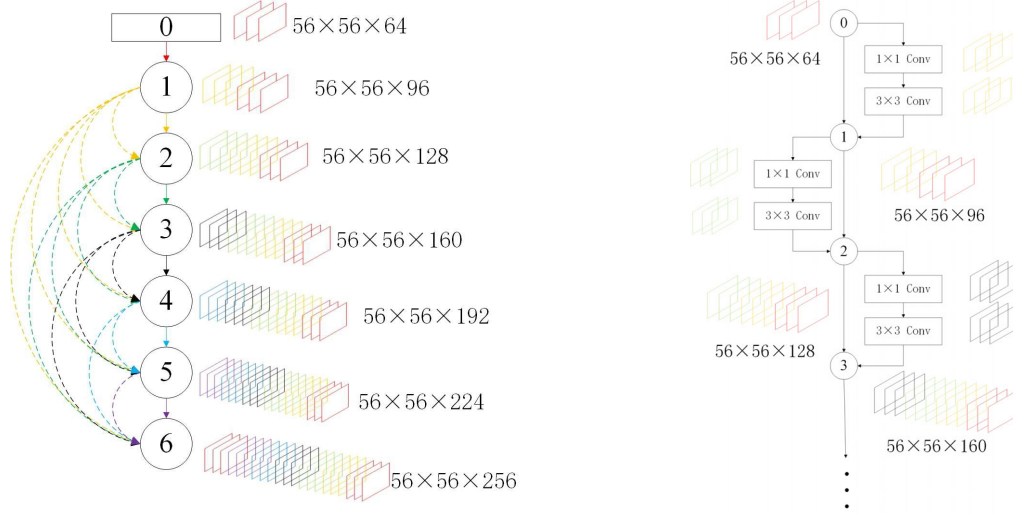
(a) Concatenated feature-maps in each layer. The solid lines denote the concatenation of previous feature-maps and the feature-maps produced in the next layer where the arrow is pointing to. The dashed lines present the connection of different layers.

(b) Newly produced featured-maps in each layer. Bottleneck layer consists of 1×1 and 3×3 convolutions.

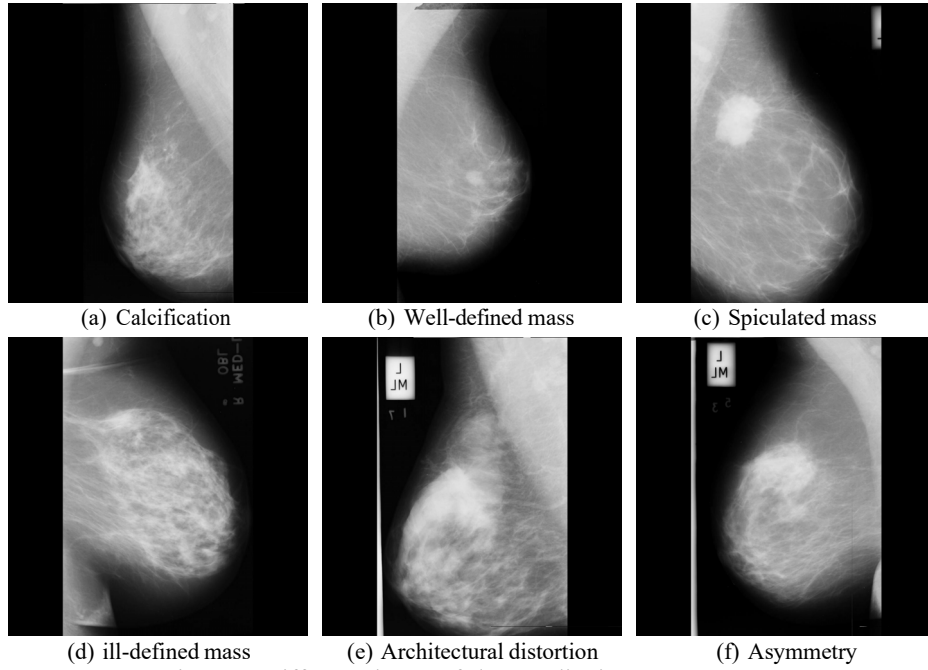Figure 1: Connections of first dense block in DenseNet201 (growth rate $k = 32$)

(a) Calcification

(b) Well-defined mass

(c) Spiculated mass

(d) ill-defined mass

(e) Architectural distortion

(f) Asymmetry

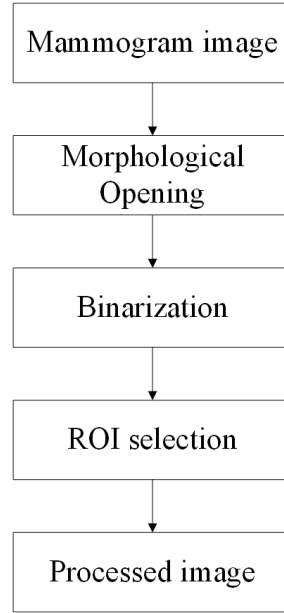Figure 2: Different classes of abnormality in mammograms

Figure 3: Preprocessing procedures

3.2   Preprocessing

As can be seen from the examples above, breasts are the foreground of images while the background is composed of the black region, white bars, and boxes. However, within the foreground, we only focus on ROIs located in breasts. Therefore, the procedure of processing can be divided into two more steps, namely the removal of background and selection of ROIs. The flow chart Fig. 3 below illustrated the procedures of acquiring ROIs from mammograms. Morphological opening, erosion followed by dilation, was first applied to remove potential noises in the image. Such that, most of the items including characters and white bars are removed. The structural element in opening operation is a disk-like filter in size of 30 pixels. To get the mask of the breast in a mammogram image, we then have the image after opening operation binarized by the threshold value, which was the average intensity of the whole image being opened. The mask of the breast was determined by selecting the connected component with the biggest area from all of the components. Fig. 4 shows the outcome of each step. According to the labels given by radiologists, we are able to determine the ROIs from the original image. Fig. 5 are corresponding ROIs to Fig. 2. Finally, we obtained 330 ROIs in total from 322 mammogram images due to the situations where one mammogram contains several suspicious regions.
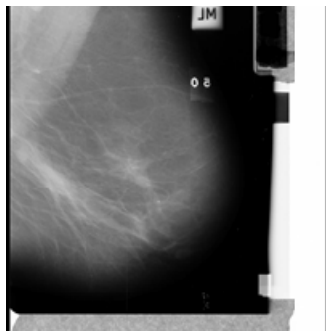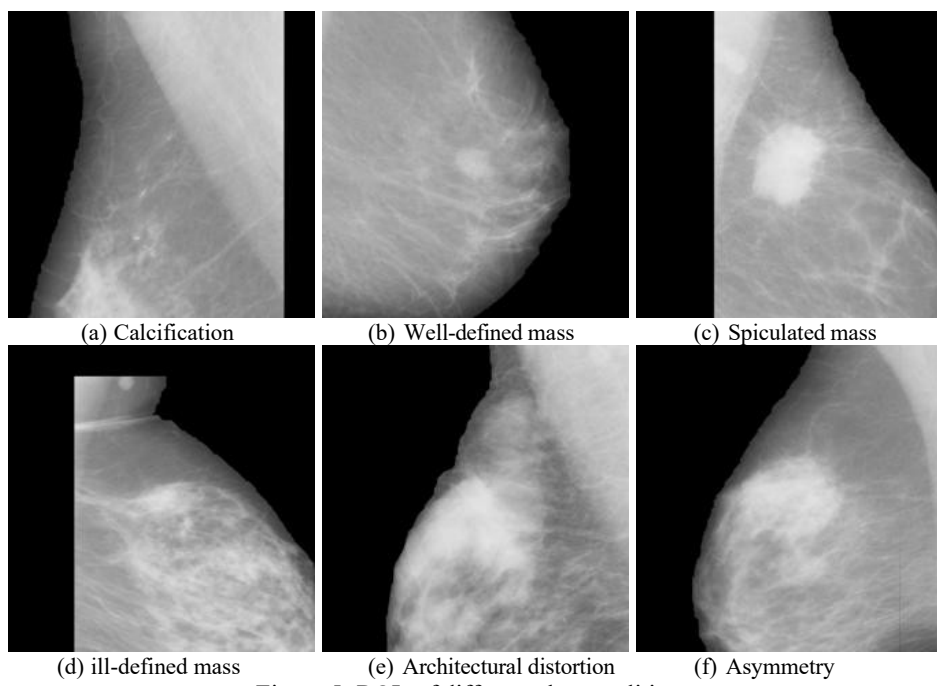

(a) Original mammogram image

(b) Opening followed by binarization
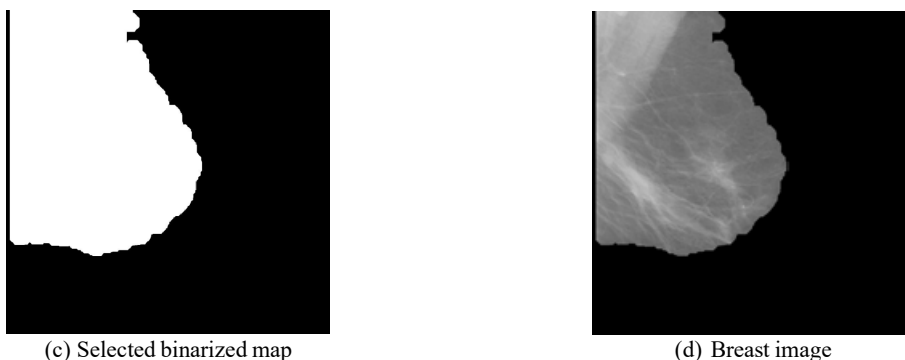
(c) Selected binarized map   (d) Breast image

Figure 4: Extraction of breast image

(a) Calcification   (b) Well-defined mass   (c) Spiculated mass

(d) ill-defined mass   (e) Architectural distortion   (f) Asymmetry

Figure 5: ROIs of different abnormalities

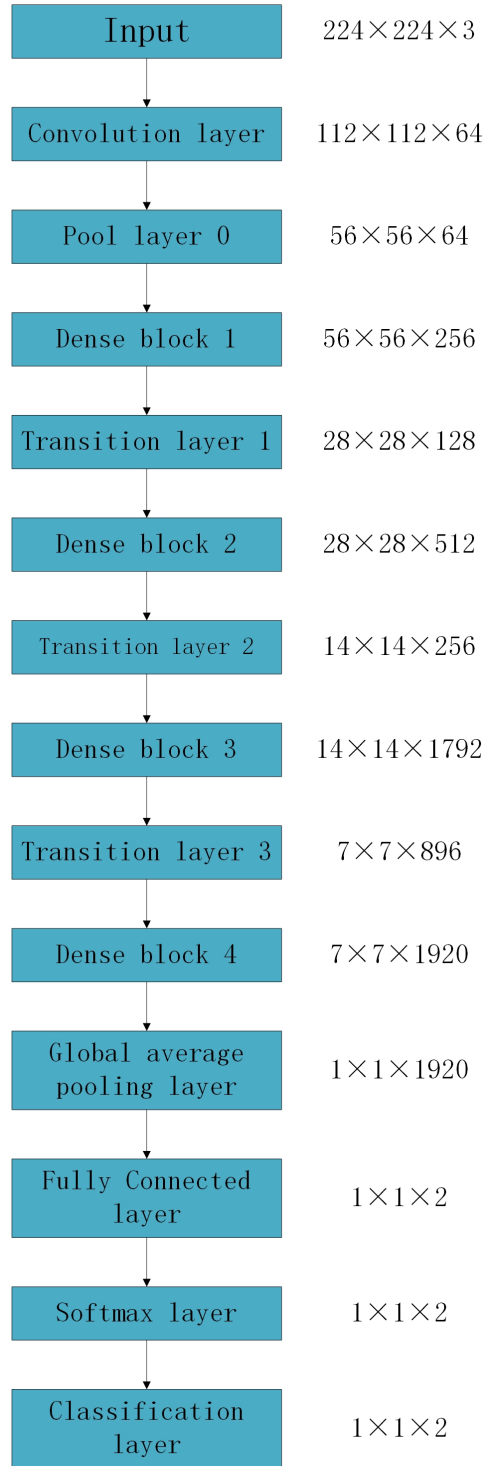|	
Input	$224 \times 224 \times 3$
Convolution layer	$112 \times 112 \times 64$
Pool layer 0	$56 \times 56 \times 64$
Dense block 1	$56 \times 56 \times 256$
Transition layer 1	$28 \times 28 \times 128$
Dense block 2	$28 \times 28 \times 512$
Transition layer 2	$14 \times 14 \times 256$
Dense block 3	$14 \times 14 \times 1792$
Transition layer 3	$7 \times 7 \times 896$
Dense block 4	$7 \times 7 \times 1920$
Global average pooling layer	$1 \times 1 \times 1920$
Fully Connected layer	$1 \times 1 \times 2$
Softmax layer	$1 \times 1 \times 2$
Classification layer	$1 \times 1 \times 2$

Figure 6: Structure of transferred DenseNet201

3.3   Transfer learning methods

Transfer learning turns out to be an effective way to introduce networks with good performance on classification into practical classification problems provided limited data size. The state-of-the-art deep CNNs are, however, generally designed for large scale objects recognition [27] [14] [21]. Therefore, in the first step of transfer learning, the fully connected layer with 1000 neurons has to be replaced correspondingly with the one with fewer neurons to meet specific classification need. Whether freezing or fine-tuning technique should be applied depends on how many parameters to be trained and the size of data. To differentiate the different performance of these two techniques, we used both of the techniques when transferring base networks because of the vagueness of the size of the data as well as the number of parameters to be trained. Considering a binary classification in our work, we replaced all of the fully connected layers in the base networks with a 2-neuron one. Therefore, only three top layers including the fully connected layer, softmax layer, and classification layer were retrained when we froze the layers preceding fully connected layer while retrained the whole networks by a small number of epochs to fine-tune parameters in the case of fine-tuning. The structure of transferred DenseNet201 is presented in Fig. 6. The output sizes of feature-maps are shown on the right-hand side of each block while the size of the input is on the top.

3.4 Hold-out validation

We used hold-out validation to examine the performance of our transferred networks. 80% of the whole data set was randomly partitioned in training set while the rest remained to be testing set. To rule out the possibility of accidental results, we ran hold-out validation for ten times repeatedly. Therefore, the performance of classifiers was assessed according to the mean classification accuracy produced in ten times running period. As for the reason why we hold 80% out of data set as the training set is that we followed the general hold-out rule, which partitioned 80% of whole data set as training set while remained data was used as the testing set.

4 Experiment

4.1 Configuration and parameters setting

This research used the SPECTRE High-Performance Computing Facility at the University of Leicester. All of the experiments are carried out on a machine with one single GPU Tesla P100 PCI-E 16GB. To compare the performance of two different transfer methods, we set the same parameters for the two methods. Without special specifications, the parameters are: dropout rate for the fully connected layer is 0.5, adaptive moment estimation is used as the optimization algorithm, mini-batch size of 64, initial learning rate as 0.01, the number of maximum epochs is 30, learning rate drops at the rate of 0.1 every 10 epochs.

4.2 Partition and augmentation of data

As aforementioned, we acquired 330 ROIs from 322 mammogram images and have them classified into two categories according to labels provided the radiologists. Then we named all of them by numbers from 1 to 330 and randomly partitioned 80% of them into the training set. Fig.7 shows the indexes of the training set and testing set. For a better understanding, we split the 1×330 array into a 5×66 array. The yellow regions indicate that the corresponding ROIs are partitioned into the training set while the blue regions denote testing set. The detailed partition of the data set is presented below in Tab 1. To avoid the problem brought by the imbalance of positive and negative samples, data augmentation was applied here. Followed the general practice, we applied random rotation from 0 to 360 and random scaling from 0.5 to 1.
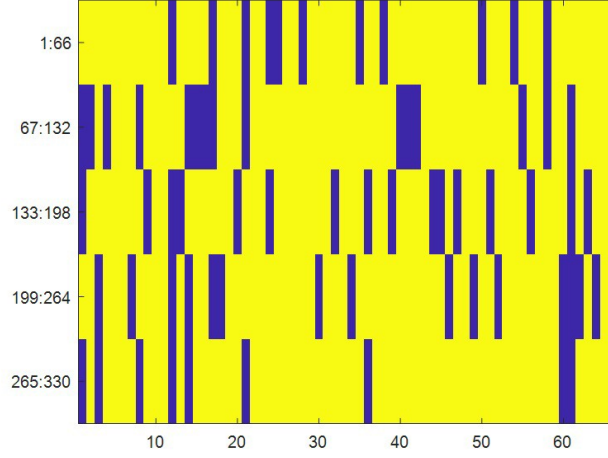
Figure 7: Partition of data

Table 1: Training set and testing set of data set

Categories	Training set	Testing set	Overall
Abnormality	98	24	112
Normality	166	42	208
Overall	264	66	330

Table 2: Comparison of two transfer methods

Model	Sensitivity (%)	Specificity (%)	Overall accuracy (%)	Mean training time (Seconds)
Freezing	86.67	75.71	79.70	**7.72**
Fine-tuning	**94.58**	**91.67**	**92.73**	1.62×10^3

4.3 Freezing or fine-tuning

It needs to be clarified that DenseNet201 here and afterward is referred to the one with last fully connected layer being replaced with a two-neuron fully connected layer. As there were not quantitative criteria when freezing or fine-tuning should be applied, we explored which method would work better in our work. We ran each configuration with the same training set and testing set for ten times repeatedly. In Tab 2, the mean accuracy of the two methods is presented. As can be seen from the table, the performance of fine-tuned DenseNet201, both on sensitivity and specificity, outperformed the one with previous layers being frozen (as referred to DenseNet201-0) by a large margin. The reason why fine-tuned DenseNet201 gained better results could be that the DenseNet-0 lost the ability to learn enough specific features from ROIs because of the fixed values of parameters. However, the expenditure on training time of the former, which reaches 1.62×10^3 seconds on average, is significantly larger than that of the latter one.

4.4 DenseNet201 being retrained in different depth

While fine-tuned DenseNet201 achieved high accuracy of classification, it is worth noting that the cost of

training is too expensive. Therefore, can we find a trade-off between accuracy and computational expenditure? Towards this aim, we retrained DenseNet201 with different depth. To avoid misunderstandings, we have DenseNet201-A, DenseNet201-B in correspondence to DenseNet201 with the last two and last one dense block being retrained. Then, we denote the DesneNet201 with all of the layers being retrained as DenseNet201-C. The detailed results are presented in Tab 3. Analysis of overall accuracy is shown in Fig 8. In terms of sensitivity, it seems the more dense blocks being retrained, the higher the accuracy is. A reasonable explanation of this could be that when previous layers in the network are able to learn some specific features while these features are lost when leaving the previous layers frozen. Interestingly, the performance of DenseNet201-A and DenseNet201-B is quite similar while training time for DenseNet201-B is 5 times more than that of DenseNet201-A's. Therefore, it seems that an appropriate depth of the base network to be frozen can lead to a significant decrease in training time without harming the accuracy too much.

Table 3: Comparison of DenseNet201 retrained in different depth

Model	Sensitivity (%)	Specificity (%)	Overall Accuracy (%)	Mean training time (Seconds)
DenseNet201-A	90.83	86.67	88.18	**105.59**
DenseNet201-B	91.25	85.71	87.73	535.11
DenseNet201-C (Our method)	**94.58**	**91.67**	**92.73**	1.62×10^3

Table 4: Accuracy of different methods

Models	Method	Accuracy (%)
Yang Li[33]	MIP + TPS	84.80 ± 3.10
Gorgel Sertbas[11]	SWT-SVM	90.10
Liu[37]	WFRFT + PCA + SVM	92.16 ± 3.60
Wu[38]	FRFE + CAR-BBO	92.52
Nguyen[35]	HMI + FNN	73.50 ± 1.35
Our method	DenseNet201-C	92.73 ± 2.84

5 Discussion

Generally, preprocessing is an indispensable procedure in traditional computer-aided diagnosis system but can be removable in systems equipped with state-of-the-art networks under certain conditions. In our system, we added the preprocessing procedure to make sure that the disturbing items in mammogram images can be removed since preprocessing is almost real-time and may contribute to higher accuracy. When it comes to transferring learning, it can be a tricky problem on whether the target networks should be fine-tuned or be frozen in previous layers. The number of parameters to be trained and the size of data are the key factors. As shown in the experiment result, the fine-tuning technique performed better than freezing in this work. A possible reason is that networks can extract more specific features after fine-tuning. However, the number of parameters to be trained in the transferred network is only 2×1920, which can be quite small compared to that of fine-tuning. That is the reason why it requires a quite long time for the network to be fine-tuned.
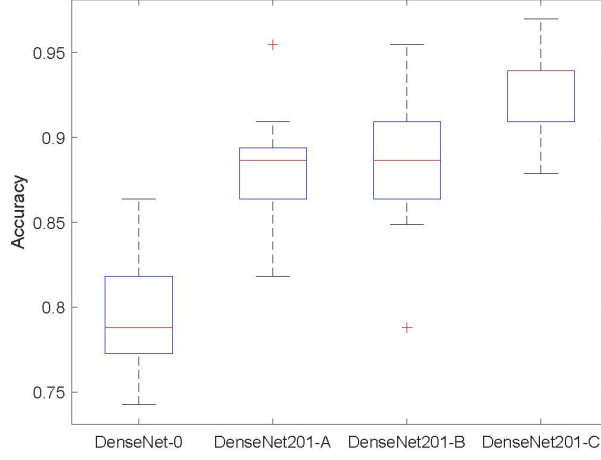
Figure 8: Accuracy of DenseNet201 trained in different depth. From DenseNet201-0, DenseNet201-A, DenseNet201-B, DenseNet201-C corresponds to DenseNet201 being retrained with only the top layer, one dense block, two dense blocks and all of the dense blocks respectively.

6　Conclusion

In this paper, we developed a semi-automatic diagnosis system for breast cancer with a mean accuracy of 92.73%. The input of the system is the original mammogram images while the output of the system is the disgnostic results including normality and abnormality, where early syndromes such as calcification, mass, and distortion in mammogram images are considered as abnormality. Also, to achieve an accuracy as high as possible, we transferred the-state-of-the-art network DenseNet201 into our system. By making a comparison of diagnostic results of networks trained by freezing and fine-tuning method, we found that the more dense blocks being retrained the higher sensitivity is. Also, the specificity and sensitivity remain high in models though transfer methods are different. High specificity and sensitivity, in turn, contribute a high accuracy, which was further enhanced by the comparison results of our method and the other state-of-the-art methods.

7　Future Prospects

Though we presented a semi-automatic diagnosis system in this paper, there are several aspects to be improved. One is that our method is only applicable to binary classification. In practical, it is of great significance to classify images to be diagnosed into normality and abnormality to improve the working efficiency of radiologists. However, classifying abnormalities into detailed categories would help radiologists with better decision making policy and improve the diagnosis accuracy of different kinds of abnormalities. Therefore, we will focus on the classification of abnormalities in our future work. Our design of semi-automation, which limits the capability of our system in practical scenarios, also needs to be improved. Therefore, we will design abnormality detection system to locate the suspicious regions in mammogram images since detection is a basic yet indispensable part in CAD systems. To solve the problem brought by the limited size of data, we will examine our method on data with a larger size. Based on these works, we are able to develop an end-to-end detection and diagnosis CAD system for breast cancer.

Acknowledgment

References

[1] Breast cancer treatment. https://www.cancer.gov/types/breast/patient/breast-treatment-pdq#section/all? redirect=true. Accessed Jan 5, 2019.

[2] Imagenet classification challenge 2012. http://image-net.org/challenges/LSVRC/2012. Accessed Dec 28, 2018.

[3] The mini-mias database of mammograms. http://peipa.essex.ac. uk/info/mias.html. Accessed Jan 4, 2019.

[4] Ayelet Akselrod-Ballin, Leonid Karlinsky, Sharon Alpert, Sharbell Hashoul, Rami Ben-Ari, and Ella Barkan. A cnn based method for automatic mass detection and classification in mammograms. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pages 1–8, 2017.

[5] Estefanía D Avalos-Rivera and Alberto de J Pastrana-Palma. Classifying microcalcifications on digital mammography using morphological descriptors and artificial neural network. In *Ciencias de la Informática y Desarrollos de Investigación (CACIDI), IEEE Congreso Argentino de*, pages 1–4. IEEE, 2016.

[6] Yusuf Aytar and Andrew Zisserman. Tabula rasa: Model transfer for object category detection. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2252–2259. IEEE, 2011.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

[8] Neeraj Dhungel, Gustavo Carneiro, and Andrew P Bradley. Automated mass detection in mammograms using cascaded deep learning and random forests. In *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*, pages 1–8. IEEE, 2015.

[9] Mengyue Geng, Yaowei Wang, Tao Xiang, and Yonghong Tian. Deep transfer learning for person re-identification. *arXiv preprint arXiv:1611.05244*, 2016.

[10] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifi er neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.

[11] Pelin Görgel, Ahmet Sertbas, and Osman Nuri Uçan. Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines. *Expert Systems*, 32(1):155–164, 2015.

[12] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22):2402–2410, 2016.

[13] William Hang, Zihua Liu, and Awni Hannun. Glimpsenet: Attentional methods for full-image mammogram diagnosis.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[15] Chen Wu, Yang Liu, Ding Feng, Fubin Wang, Paul Tu. Femtosecond laser ablation power level identification based on the ablated spot image. The International Journal of Advanced Manufacturing Technology, 94(5-8):2605–2612, 2018.

[16] Liu Yang, Ding Feng, Paul Tu, Fubin Wang, Chen Wu. Pseudo-color enhancement and its segmentation for femtosecond laser spot image. Microwave and Optical Technology Letters, 60(4):854–865, 2018.

[17] Fubin Wang, Ao Xu, Kai Zeng, Zhikun Chen, Yaluo, Zhou. Research for billet limited weight cutting based on behavior stateflow. In MATEC Web of Conferences, volume 68, page 02005. EDP Sciences, 2016

[18] Shuihua Wang, Yi Chen. Fruit category classification via an eight-layer convolutional neural network with parametric rectified linear unit and dropout technique. Multimedia Tools and Applications, pages 1–17, 2018.

[19] Yi Chen, Yin Zhang, Ming Yang, Huimin Lu, Hainan Wang, Bin Liu, Preetha Phillips, Shuihua Wang, Tianming Zhan. Multiple sclerosis detection based on biorthogonal wavelet transform, rbf kernel principal component analysis, and logistic regression. IEEE Access, 4:7567–7576, 2016.

[20] Kai-Lung Hua, Che-Hao Hsu, Shintami Chusnul Hidayati, Wen-Huang Cheng, and Yu-Jen Chen. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoTargets and therapy*, 8, 2015.

[21] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *CVPR*, volume 1, page 3, 2017.

[22] Benjamin Q Huynh, Hui Li, and Maryellen L Giger. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3):034501, 2016.

[23] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[24] Afsaneh Jalalian, Syamsiah BT Mashohor, Hajjah Rozi Mahmud, M Iqbal B Saripan, Abdul Rahman B Ramli, and Babak Karasfi. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clinical imaging*, 37(3):420–426, 2013.

[25] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.

[26] Jan-Jurre Mordang, Tim Janssen, Alessandro Bria, Thijs Kooi, Albert Gubern-Mérida, and Nico Karssemeijer. Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. In *International Workshop on Digital Mammography*, pages 35–42. Springer, 2016.

[27] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bern- stein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[28] Ravi K Samala, Heang-Ping Chan, Lubomir Hadjiiski, Mark A Helvie, Jun Wei, and Kenny Cha. Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. *Medical physics*, 43(12):6654–6666, 2016.

[29] Ling Shao, Fan Zhu, and Xuelong Li. Transfer learning for visual categorization: A survey. *IEEE transactions on neural networks and learning systems*, 26(5):1019–1034, 2015.

[30] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[31] Ravipudi Venkata Rao, and Peng Chen. Abnormal breast detection in mammogram images by feed-forward neural network trained by Jaya algorithm. *Fundamenta Informaticae*, 151(1-4):191–211, 2017.

[32] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big Data*, 3(1):9, 2016.

[33] Shih-Neng Yang, Fang-Jing Li, Yen-Hsiu Liao, Yueh-Sheng Chen, Wu- Chung Shen, and Tzung-Chi Huang. Identification of breast cancer using integrated information from mri and mammography. *PloS one*, 10(6):e0128404, 2015.

[34] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014.

[35] Xiaowei Zhang, Jiquan Yang, and Elijah Nguyen. Breast cancer detection via hu moment invariant and feedforward neural network. In *AIP Conference Proceedings*, volume 1954. AIP Publishing, 2018.

[36] Chichun Pan, Xianqing Chen, and Fubin Wang. Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling. *Journal of Computational Science*, 27:57–68, 2018.

[37] Ge Liu, and Jiquan Yang. Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform. *Advances in Mechanical Engineering*, 8(2), 2016.

[38] Xueyan Wu, Siyuan Lu, Hainan Wang, Preetha Phillips, and Shuihua Wang. Smart detection on abnormal breasts in digital mammography based on contrast-limited adaptive histogram equalization and chaotic adaptive real-coded biogeography-based optimization. *Simulation*, 92(9):873-885, 2016.