2019 10th International Conference on Information Technology in Medicine and Education (ITME)

Deep Learning from Small Dataset for BI-RADS Density Classification of Mammography Images

Peng Shi*, Chongshu Wu
College of Mathematics and Informatics
Fujian Normal University
Fuzhou Fujian China
*pshi@fjnu.edu.cn

Jing Zhong
Department of Radiology
Fujian Cancer Hospital & Fujian Medical University Cancer Hospital
Fuzhou Fujian China

Hui Wang
School of Computing and Mathematics
Ulster University
Jordanstown Antrim UK

ABSTRACT

Mammography is a breast imaging technique that has been widely used in breast cancer diagnosis and screening. The Breast Imaging Reporting and Data System (BI-RADS) defines a six-point overall cancer risk scale from negative to highly suggestive of malignancy based on mammography, and also a four-point breast density based cancer risk scale. Automatic BI-RADS density classification of mammogram images is still a challenge. The current state of the art is about 80% on the MIAS (Mammogram Image Analysis Society) database. In this paper we present a deep learning study of BI-RADS density classification using MIAS, based on a lightweight Convolutional Neural Networks (CNNs) architecture. This is a small data problem as MIAS has only 322 images with ground truth, so we use image pre-processing and augmentation to solve the problem. Five-fold cross validation is used to evaluate the proposed approach, and has achieved a test accuracy of 83.6% on average. This suggests that deep learning has the potential to address the small data problem in mammography, which is prevalent in many medical image analysis tasks. The experience we have, especially in how to optimize the deep learning architecture, will benefit other researchers and medical practitioners.

KEYWORDS

Deep learning, Convolutional neural networks, Breast cancer, Mammography, Breast density classification

1 Introduction

Breast cancer is one of the most common cancers among women [1]. It has been shown that early diagnosis is associated with reduced breast cancer morbidity and mortality [2], therefore early diagnosis of breast cancer has become a focus of attention in recent years, especially in the medical and computer science research communities. Imaging techniques including ultrasonography [3], Computed Tomography (CT) [4], Magnetic Resonance Imaging (MRI) [5] and mammography [6] have been widely used in breast cancer diagnosis and screening. Among them, mammography screening plays an important role due to the practice being able to reduce breast cancer risks significantly and especially benefiting women aged 40 to 49 years [7].

Mammography generates high-resolution mammogram images with low-energy X-rays, and provides the possibility of discovering abnormalities masked by surrounding and overlapping breast tissue [8]. In clinics, two kinds of mammography are being used including analog Screen-Film Mammography (SFM) and Full-Field Digital Mammography (FFDM) [9]. Due to the need of digital mammogram images for automatic analysis, SFM images are also being digitally scanned in Computer Aided Diagnosis (CAD) systems, providing clinical evidence for human experts. A few databases of digital mammogram images have been made public for research purposes, such as Mammogram Image Analysis Society (MIAS) [10], Digital Database for Screening Mammography (DDSM) [11], and INBreast [12]. As one of the earlier released databases, MIAS has been widely used in breast cancer related research including breast segmentation, breast tissue classification, micro-calcification detection and mass classification. MIAS provides meta data about the mammogram images including characteristics of background tissue, classification of abnormality, severity of abnormality, locations of micro-calcification and ground truths of breast region for segmentation.

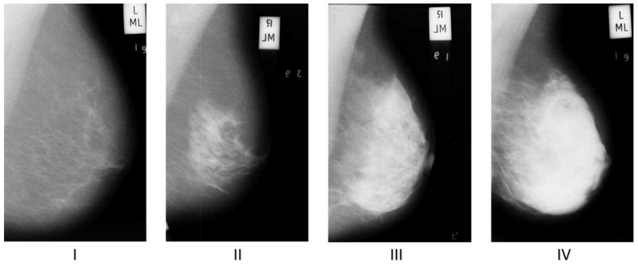

Figure 1: Images from MIAS database with different breast densities in BI-RADS density classification

Among multiple features drawn from a mammogram image, density is highly related to breast cancer risk [13]. In Breast Imaging Report and Data System (BI-RADS) proposed by American College of Radiology (ACR), a six-point scale (BI-RADS-6) is defined to classify the malignancy of breast cancer [14]. Similarly, a four-point scale (1-4) is defined to classify breast density, which includes (a) predominantly fat, (b) fat with some fibroglandular tissue, (c) heterogeneously dense and (d) extremely dense [15]. Figure 1 shows sample mammogram

978-1-7281-3918-0/19/$31.00 ©2019 IEEE
DOI 10.1109/ITME.2019.00034

images from MIAS with different densities classified in BI-RADS. Automatically classifying breast densities using the four-point scale is highly desirable, yet it is still a challenge.

Deep learning techniques, such as Convolutional Neural Network (CNN), have been very successful in a broad range of applications, especially in image comprehension and classification [16]. However, conventional CNNs need big data to work well since they usually have a huge number of parameters to optimize through training. This makes them unsuitable for applications that do not have large datasets such as medical applications.

In this paper, we propose a lightweight CNN-based deep learning framework for BI-RADS density classification of mammogram images in MIAS that addresses the small data challenge in deep learning as well as the efficiency challenge. We achieve five-fold cross validation classification accuracy of 83.6% on average, which is above most of the current methods. This suggests that deep learning has the potential to overcome the small data challenge, which is prevalent in medical image analysis applications. The rest of paper is organized as follows. In Section 2, we review existing work in the literature on BI-RADS density classification and deep learning on small datasets. In Section 3 we present technical details of our proposed method. In Section 4 we present experimental results with detailed quantitative evaluation and comparisons, along with discussions. Finally in Section 5 we present conclusions and future work.

2 Introduction BI-RADS density classification and deep learning: a literature review

2.1 BI-RADS Density Classification

There are broadly two approaches to mammogram image based breast density classification -- feature-based and machine learning-based. Due to the complexity of tissue appearance in the mammograms such as a wide variation and obscure texture patterns within the breast region, most of reported feature-based methods have accuracies lower than 80% [17-19], in which first and second-order texture features such as Local Binary Patterns (LBP), Local Grey-level Appearance (LGA), and Basic Image Features (BIF) are tested, and obtained accuracy of 75% as the highest and only one accuracy exceeding 90% is presented by Parthalain et al. [20] using a sophisticated feature selection framework.

Machine learning-based methods are generally used in lesion classifications of mammography, which are generally based on calculations of distances between feature vectors extracted from different images. Fischer et al. [21] propose a Bayesian networks to classify breast lesion under BI-RADS standards. Raúl et al. [22] build a set of machine learning classifiers for breast cancer BI-RADS diagnosis, and a hierarchical similarity measurement scheme based on a distance weighting function is proposed by Wei et al. [23] to search for similar mammograms in the database. Above works have shown good performances suing machine learning approaches in mammography analysis, but feature design is still an essential step in those conventional machine learning-based approaches for distance calculations.

With the rapid development of deep learning, a number of studies concerning mass classification based on deep learning are provided. Arevalo et al. [24] apply CNN for mammography mass lesion classification. Qiu et al. [25] propose an eight layer deep learning network for automatic feature extraction and a multiple layer perception classifier for feature categorization. While Carneiro et al. [26] apply a pre-trained network from the general image database of ImageNet [16] for the density classification on mammography image databases including DDSM and INBreast, and the migration learning has archived over 90% of accuracy in for a two-class problem - benign and malignant, indicating a new comprehensive way of addressing this challenging classification problem by deep learning.

2.2 Deep Learning on Small Dataset

Unlike databases containing thousands of mammography images, the scale of MIAS database is much smaller with hundreds of images, which is also the common problem of public released medical image databases with ground truths. In deep learning framework, CNNs enable learning data-driven, highly representative, hierarchical image features from sufficient training data [27]. Therefore, employ deep learning for classifications of small datasets with insufficient training data samples, image augmentation is highly demanded to extend training data with containing similar features from the original database.

Image augmentation is the process of taking images that are already in a training dataset and manipulating them to create many altered versions of the same image. It generate batches of image data with real-time data augmentation based on the limited inputs, and the data will be looped over in batches for training the network, which also makes the network robust to image variations.

Image enhancement and deformation are two main strategies of image augmentation for deep learning network. Image enhancement always include normalization and whitening transformation to change the light and color of input images, while image deformation applies flip, rotation, shift and zoom to expand the image dataset within certain reasonable degrees. The indefinite images generated by augmentation help the deep learning network work properly on small dataset.

3 Materials and Methods

To deal with mammography density classification more efficiently, a deep learning framework based on CNN is proposed as shown in Figure 2. Five-fold cross validation is applied to divide the input datasets randomly into two parts, in which 8/10 of images are used for training, and the rest 2/10 of images compose the test dataset to validate the performance of the trained network. Accuracies and losses of different strategies are compared to find the optimized combination of image segmentation, image augmentation and CNN framework for BI-RADS density classification of mammography images.

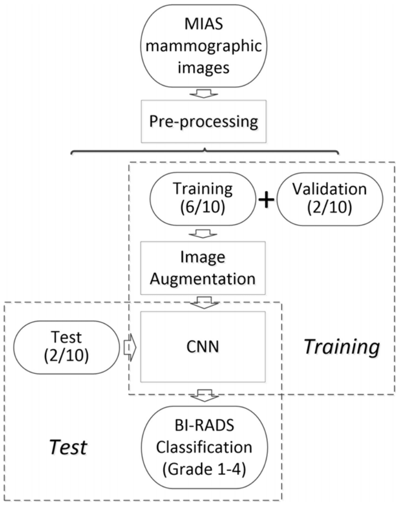

Figure 2: Deep learning strategy for MIAS mammography classification

3.1 MIAS Database

As one of the first publicly released mammography datasets, the MIAS database contains the original images of digitalized SFM at 50 micron resolution [10], and is associated truth data of breast boundary, character of background tissue and especially the locations of calcifications and various masses, in which 8-bits gray scale images are obtained from the original data format, representing the optical density of tissues.

To validate the proposed algorithm, we use the MIAS dataset v1.21 including 322 mammography images, and unify different original resolutions to 200×200 in pixels as inputs of the following deep learning framework, where resizing keep most of useful information of images and sharply decrease the need of memory in the network training.

3.2 Image Segmentation

Since abnormal tissues including some ill-defined masses or calcification always appear inside the breast region, a key part of mammography analysis is image segmentation which estimates skin-air boundary and pectoral-breast boundary which together define the breast contours. As shown in Figure 3, most of digitalized SFM contain artifacts such as labels, markers, scratches, and even adhesive tapes, which also need to be removed for image analysis. However, image segmentation may be time consuming for deep learning classification. The affects of image segmentation need to be compared before and after breast segmentation to find the best strategy of mammography density classification.

Figure 3 shows two different segmentation strategies which detect skin-air boundary and breast region boundary respectively. The difference is whether the pectoral muscle is kept or not. To get the skin-air boundary, we firstly calculate the weight for each pixel based on the gradient magnitude at that pixel, and returns a weight array. The weight of a pixel is inversely related to the gradient magnitude combining gradients of both vertical and horizontal direction at the pixel location, which separated higher intensity areas from low intensity background because sharp gradient magnitude changes occurred on the edges between foreground and background. We also apply the ground truth of breast regions provided by MIAS to get pure breast region for comparison.

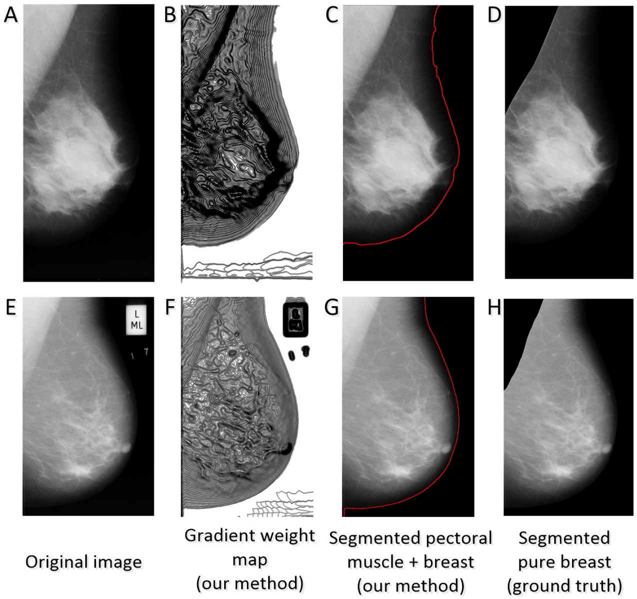
A
B
C
D
E
L ML
F
G
H
Original image
Gradient weight map (our method)
Segmented pectoral muscle + breast (our method)
Segmented pure breast (ground truth)

Figure 3: Images segmentation for skin-air boundary and breast region boundary respectively, in which (A) and (E) are original digitalized SFM images, (B) and (F) are gradient maps for skin-air boundary segmentation based on our method, (C) and (G) are segmented pectoral muscle and breast regions marked with thin red lines against the background, and (D) and (H) are segmented pure breast regions based on ground truth from MIAS

3.3 Image Augmentation

3.3.1 Image Enhancement Based on Whitening Transformation

Due to different illuminations of background in mammographic imaging, a whitening transform of an image is necessary which reduces the redundancy in the matrix of pixel images. Less redundancy in the image is intended to better highlight the structures and features in the image to the learning algorithm. Zero-phase Component Analysis (ZCA) [28] results in transformed images that keeps all of the original dimensions and the resulting transformed images still look like their originals, and it is used in [29] to correct intensity to reduce such brightness attenuation. A convolutional 5×5 ZCA whitening transform is applied in our algorithm to remove first-order correlations between neighboring pixels in the gray level image, which generates more training image samples for the following CNNs with smaller brightness differences than the original ones.

3.3.2 Image deformations

To further expand the size of input dataset, linear deformations including flip, rotation, shift and zoom are employed. All the ranges of deformation are set as 0.2 for balance between accuracy and robustness of the network. In each batch of training, 32 deformed images are generated based on the original MIAS image inputs, which solved the small data problem well in the application of CNN network

3.4 Lightweight CNN

In the large scale image classification, most of the CNN architectures of deep learning are full weight with many CNN layers, such as VggNet [30] and ResNet [31] for ImageNet competition. The number of trainable parameters in a full weight CNN is quite huge, for example, above 15 million for VGG16 [30], which needs big data and calculation resources to be trained. To fit for the small dataset, the layers of CNN in our network are greatly reduced to the combination of two or three convolutional layers. The training of the lightweight CNN is much more efficient with a good performance in practice.

3.4.*1* Convolutional Layers

A convolutional layer contains multiple kernels of weight matrices which extract certain features by padding on the input image, and calculate multiplication of the weight matrix and the highlighted part of the input image. To simplify the structure of CNN, the kernels of all convolutional layers are set as 3×3 matrices in our framework, which are used for producing the feature maps. The Rectifier Linear Unit (ReLU) [32] activation function is applied in each CNN layer as defined in Eq. 1.

$$f(x) = \max(x, 0) \tag{1}$$

where the ReLU reduces likelihood of vanishing gradient than Sigmoid activation function, which also accelerates the training of CNN in practice.

3.4.*2* Light weight Deep Learning Frame

Besides convolutional layers, other types of layers are also included to form the network. Max pooling layers of 2×2 matrices downscale the input matrix size with taking the maximum value of each window. Dropout layers consist in randomly setting a fraction rate of input units to 0 at each update during training time, which helps prevent overfitting. Dense layers regularize the densely-connected layer and the Softmax function as shown below is used in the output layer to classify input images into four classes.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{2}$$

where j indexes the output units (j = 1, 2, ..., K) to a K-dimensional vector z of real values in the range [0, 1] from the inputs to the output layer. The structure of the lightweight CNNs is shown in Figure 4.

Figure 4 shows a lightweight CNN structure is composed by three convolutional layers and three connecting layers. Input images are reshaped as 200×200 in size at the first step. Then, 32, 32 and 64 convolutional cores with size of 3×3 pixels are used in each layer respectively to extract deep features of input mammography images. Finally, output of four BI-RADS density classes are driven after a set of connecting layers which convert 2D array to 1D for classification calculations. Comparing to VggNet of sequential CNNs model, the total parameters in the above network is much smaller than that of full weight network such as VGG16, which has 16 weight layers and much more convolutional cores. It indicates the training could be time and computing saving in practical use.

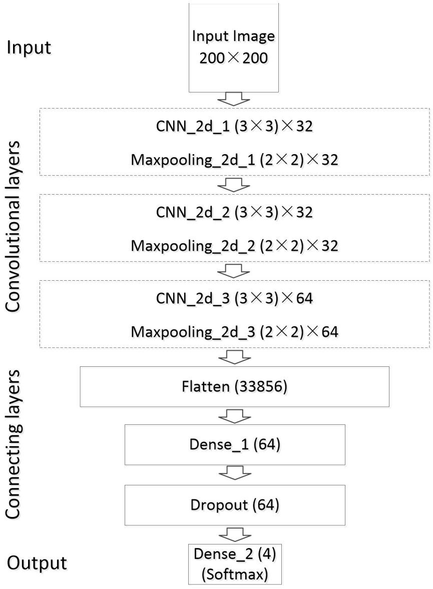

Figure 4: Deep learning network with three convolutional layers

4 Experimental results

To find the best image processing strategy and deep learning network structure, three sets of experiments are designed – (1) with and without image segmentation (2) with and without augmentation (3) lightweight CNN and full weight CNN. For image pre-processing, we test the impact of segmentation and augmentation on classification results. We also further reduce the scale of CNNs by removing few layers to find the balance between classification accuracy and training efficiency of CNN combinations. Image pre-processing related tasks are performed in MATLAB. All matrix operations for deep learning are performed with Keras with Tensorflow as backend on a laptop-based NVidia GeForce GPU with 2GB video memory, showing the lightweight network is capable to run in a low configuration environment. During training, 200 epochs are generally needed to get the network parameters stable. A batch of 32 images are fed as inputs in one step, and a total of 100 steps are carried out in each epoch to ensure full training of the network in limited times. Five-fold cross validation is used, and each input image is identified by one of the four classes according to the BI-RADS density categories provided by MIAS database.

4.1 MIAS Database BI-RADS Density Classification with and without Image Segmentation

To test the robustness of the proposed CNN structure, comparisons between segmented and original images are have been made. Three types of image are tested as inputs to compare the classification performance, including the un-segmented original images, segmented images including pectoral muscle and breast regions, and segmented images of pure breast regions based on ground truth. After 200 epochs, both validation and training performances become stable with parameters of convolutional layers being well trained. The training and validation accuracies/losses on different datasets are shown in Figure 5 and Table 1.

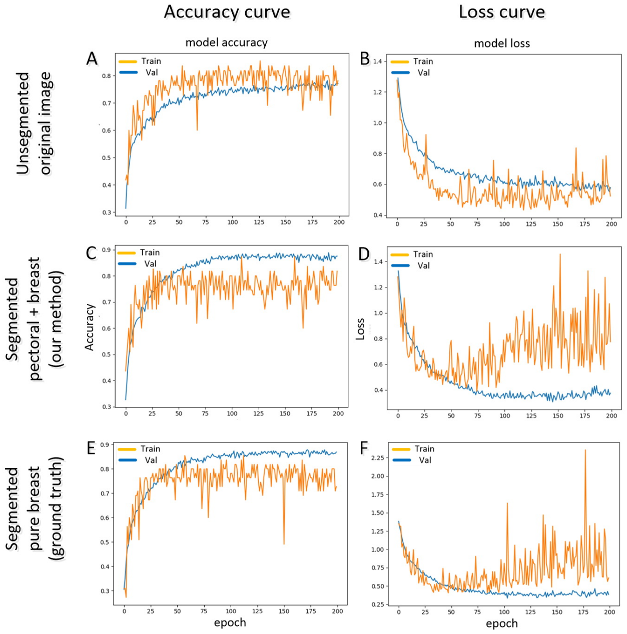

Figure 5: Accuracy (A, C, E) and loss (B, D, F) curves of original and segmented datasets respectively, where x axis is the epoch number during training and validation, y axes are classification accuracies and losses respectively, the amber lines are curves of validation, and the blue lines are curves of training

As shown in Figure 5, generally major variations always appear in training curves (amber lines), which make them have great change amplitude, and the smooth validation curves (blue lines) mainly show the trends of accuracy increasing based on deep learning. According to the validation curves, the unsegmented images from original dataset has the lowest accuracy of 78.2%, suggesting artificial like labels and tapes may affect the image classification performance of CNN. Meanwhile, more focused ROIs including segmented pectoral muscle and breast, or the pure breast region archive higher classification accuracies, and only a little difference can be found between the last two datasets, which indicates the proposed network is not sensitive to image segmentation.

Table 1. Classification performance comparisons between original and segmented datasets.

Dataset	Original	Segmented pectoral muscle + breast	Segmented pure breast
Loss	0.73	0.52	0.80
Accuracy (%)	78.2	83.6	81.9

4.2 Heading Level 2

The improvement of data extension in small dataset is tested by comparing classification accuracies with and without image augmentation. Under the default mode of the network, image augmentation of ZCA whitening and linear shape deformations are included first in generating image batches for training. However, in the framework without image augmentations, 32 input from MIAS dataset are randomly selected from 322 images segmented images, and fed into the network. This process is repeated in each batch of training.

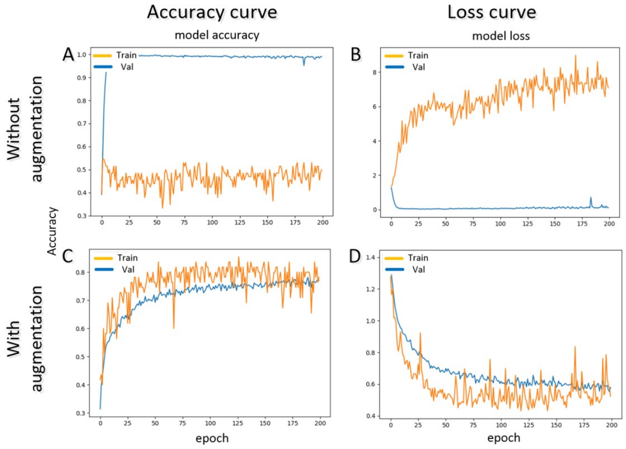

Figure 6: Accuracy and loss curves on datasets with and without image augmentation, where the amber lines are curves of validation and the blue lines are curves of training

In the case without augmentation, the training loss (Figure 6B) decreases sharply to the bottom and keeps almost a straight line after about 20 epochs, and the training accuracy (Figure 6A) quickly increases to above 99% whereas and the validation accuracy stays low with more epochs. In the case with augmentation, the validation accuracy (Figure 6C) has risen from 51.3% to 83.6%. The augmentation steps generate more related images as input rather than the originals, which expand the size of sample dataset and make the network parameters get more training. That sharply increases the classification performances for the small scale database as Figure 6 shows.

4.3 Lightweight CNNs vs. full weight CNNs

To test the performance of CNNs with fewer convolutional layers, two kinds of lightweight CNN structures are compared with the full weight VGG16 network respectively, including a 2-layer and a 3-layer CNN. Meanwhile, we also use the concept of transfer learning, in which using a well-trained VGG16 network directly on the small scale MIAS dataset for a comprehensive comparison.

4.3.1 High efficiency in training and testing based on lightweight CNN

Firstly, we compare the parameter scales and training times of each network as shown in Table 2. Both 2-layer and 3-layer CNNs have about two million trainable parameters, whereas the VGG16 network has more convolutional layers and even more convolutional cores, hence it has one order of magnitude more trainable parameters. Fast-growing parameter scales make the training of full weight network much more time consuming than lightweight networks. In the homogeneous comparison of training times per epoch, light weight networks have seven times fewer time consumption than VGG16, which highly improves the efficiency in the application of deep learning scheme.

Table 2. Comparison of network performances between different CNNs.

Number of CNN layers	2-layer	3-layer	16-layer (VGG16)
Total parameters	2,364,260	2,195,172	15,304,004
Training time per epoch (s)	37	41	285
Loss	0.60	0.52	11.79
Accuracy (%)	78.2	83.9	27.3

4.3.2 High accuracy in training and testing based on 3-layer CNN

A convolutional layer contains multiple kernels of weight matrices which extract certain features by padding on the input image, and calculate multiplication of the weight matrix and the highlighted part of the input image. To simplify the structure of CNN, the kernels of all convolutional layers are set as 3×3 matrices in our framework, which are used for producing the feature maps. Then, the Rectifier Linear Unit (ReLU) activation function is applied in each CNN layer as defined in Eq. 1.

With greatly improvement of efficiency, the performance of deep learning network mainly lies in the precise classification of samples from different categories. After segmentation to get the skin-air boundaries of input images, the above three types of networks are tested to compare their classification accuracies and losses as shown in Figure 7.

The lightweight networks become stable after 200 training epochs, while both loss and accuracy curves of VGG16 are nearly straight lines, suggesting the training is far from enough to get all the convolutional cores well trained. The performances of different CNNs are summarized in Table 2, including parameter scales, time consumption, and classification results, which gives a comprehensive view of each network. The poorly trained VGG16 based on the small dataset is excluded at first because of its bad performance. The average accuracy of 3-layer CNN is about 3% higher than that of 2-layer network, with a little increase in training time which could be ignored in practical applications.

4.3.3 High accuracy based on training the proposed 3-layer CNN than transfer learning

In order to test the performance of transfer learning, a VGG16 network with well-trained parameters based on ImageNet competition is applied for MIAS image classification. To do the transfer learning, we reserve all the parameters of the convolutional parts, and only leave the fully-connected layers to be trained. We then run this model on our training and validation datasets, and only a small amount of parameters in the fully-connected layers are trained according to the new MIAS dataset. Using the same cross validation strategy, the best accuracy of classification is 58.3%, indicating even the heavy weight CNN cannot handle the small scale dataset well. Meanwhile, the proposed light-weight structure still has better performance than transferring a common heavy-weighted network like VGG.

5 Conclusions and Discussions

Comparing to most of the existing mammogram image processing methods, the proposed method executes multiple tasks in a hierarchical way to acquire good performance in each step, which deals with specific morphologic characteristics of different objects for detection. Compared with some general purpose image databases such as ImageNet, medical image databases are usually much smaller because of the difficulties in image acquisition, identification, categorization and privacy protection. It is therefore a challenge to apply deep learning on small medical image datasets. Our strategy is to use smaller network and apply image augmentation to enlarge the image dataset. In this paper, we present a deep learning framework for BI-RADS density classification, in which a 3-layer lightweight convolutional network is employed to solve the small data problem and to improve the classification accuracy. The original 322 images in MIAS are firstly segmented using the gradient weighted map to get the foreground including pectoral muscle and breast regions as inputs to CNN. Then, image augmentation is applied to generate training data samples similar to the original images. Finally, a 3-layer CNN is trained to classify images into four BI-RADS density categories, which can be used for mammography based breast cancer screening or diagnosis.

5.1 The new state of art in BI-RADS density classification

Most of feature-based researches achieve classification accuracies below 80%, with sophisticated image processing techniques for feature set extractions and comparisons. Using our deep learning based approach, with little pre-processing, we achieved a classification accuracy of 83.6% on average. This suggests that a proper combination of image pre-processing, image augmentation and balanced CNN structure is able to solve the small data problem to some extent, which is prevalent in medical image databases in most of clinical practices.

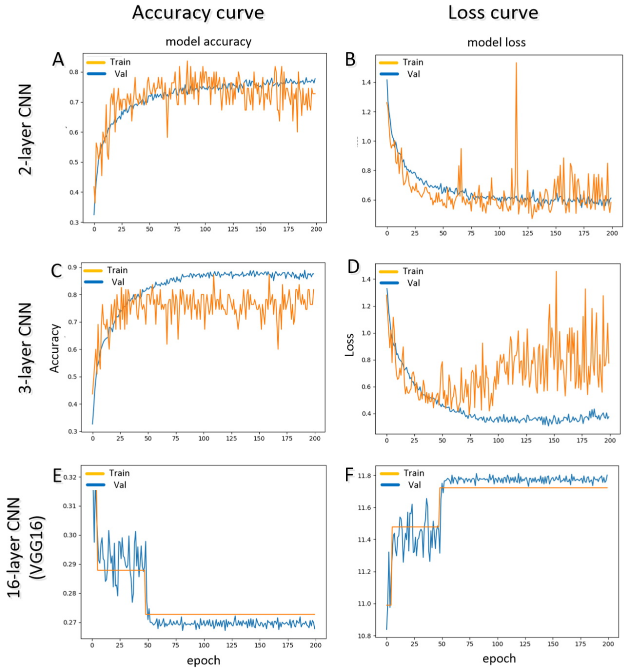

Figure 7: Accuracy (A, B, C) and loss (B, D, F) curves of different CNNs, where the amber lines are curves of validation and the blue lines are curves of training

5.2 Segmentation has little effect on classification

In medical image analysis, image segmentation is usually a key step to precisely locate the Region of Interest (ROI) and the subsequent quantifications. However, since CNN mainly works on the deep features rather than only the morphology of input images, image segmentation plays a minor role in the proposed classification method. Experimental results show that the improvement is not too much after using image segmentation, and the best performance is achieved in classifying segmented images with pectoral muscle plus breast regions rather than only the breast regions. The main reason is that pectoral muscle plus breast regions exclude the disturbance of artificial objects such as tapes and labels, and retains more image details than the smaller breast regions. This also suggests that less efforts could be put on image segmentation in applying deep learning, which may be time saving and more capable in clinical use.

5.3 Image augmentation is necessary to classification

Image augmentation is essential in enlarging training dataset for deep learning to work well. The augmentation of input images satisfies the requirements of deep learning rather than simply using original images repeatedly. Comparison experiments suggest that classification accuracy is highly improved by applying ZCA whitening and a set of linear deformations to the original images.

5.4 Limitation and future work

The main problem of deep learning on small dataset is the stability of the network. The variation ranges of the loss and accuracy curves are still relatively large, and big variations may happen occasionally. Even using image augmentation to enlarge the training dataset, it is still challenging to get all the convolutional cores well trained because of little diversity between the generated images and small amount of original samples. Experimental results also show that full weight CNNs are not capable to be directly used on small datasets.

Basically there are two more ways to solve the small data problem in deep learning, including the Generative Adversarial Network (GAN) [33] and transfer learning [34]. Using a generative model and a discriminative model, GAN generates samples from the training data with more probability of differences. GAN has begun to be used in image synthesis, including generating highly compelling images of specific categories [36], and medical image segmentation [36]. Lacking of labeled training sample in most of medical image datasets, which is a good option to expand training sample sizes for datasets like MIAS.

Applying transfer learning on small dataset is making use of pre-trained network from the general image databases with big data. Experimental results of directly transferring a network dealing with general image classification problem like ImageNet is not capable of classifying mammographic images, indicating a more specific network trained by a large scale mammogram database is needed. A pre-trained VGG16 network from ImageNet competition has been used in classification of DDSM, in which the weights of CNNs are pre-loaded and need little adjustment in working on the new dataset [27]. Therefore, transfer learning from a big scale medical image database such as DDSM to smaller database MIAS is also workable. Similar image properties between different mammography image databases may bring better performances than using general image databases.

In the future work, we will focus on the above two aspects to further improve the robustness and efficiency of CNNs on classification of mammography dataset, and explore ways of solving the small-data problem restricting the development of deep learning applications on medical image databases.

ACKNOWLEDGMENTS

This work is a joint research project undertaken by Fujian Normal University and Ulster University. The authors would like to thank every laboratory member involved in this cooperation. The project is financially sponsored by the National Science Foundation of China (61501121), Fujian Provincial Department of Science and Technology (2015J05145 and 2016Y0018), and funding from Fujian Engineering Technology Center of Big Data Mining and Applications for Public Service.

REFERENCES

[1] DeSantis, C., Ma, J., Bryan, L., et al. (2014). Breast cancer statistics, 2013. CA: A cancer journal for clinicians, 64(1), 52-62.

[2] Oeffinger, K C., Fontham, E. T. H., Etzioni, R., et al. (2015). Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA, 314(15), 1599-1614.

[3] Jalalian, A., Mashohor, S.B.T., Mahmud, H.R., et al. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clinical Imaging, 37(3), 420-426.

[4] Chen, B., and Ning, R. (2002). Cone ‑ beam volume CT breast imaging: Feasibility study. Medical Physics, 29(5), 755-770.

[5] Mann, R.M., Kuhl, C.K., Kinkel, K., et al. (2008). Breast MRI: guidelines from the European Society of Breast Imaging. European Radiology 18(7): 1307-1318.

[6] Olsen, O., and Gøtzsche, P.C. (2001). Cochrane review on screening for breast cancer with mammography. The Lancet, 358(9290), 1340-1342.

[7] Smart, Charles R., R. Edward Hendrick, James H. Rutledge, and Robert A. Smith. "Benefit of mammography screening in women ages 40 to 49 years. Current evidence from randomized controlled trials." Cancer 75, no. 7 (1995): 1619-1626.

[8] Johns, P.C., and Yaffe, M.J. (1987). X-ray characterization of normal and neoplastic breast tissues. Physics in Medicine and Biology, 32(6), 675-695.

[9] Mustra, M., Grgic, M., and Rangayyan R.M. (2016). Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. Medical & Biological Engineering & Computing, 54(7): 1003-1024.

[10] Suckling, J., Parker, J., Dance, D., et al. (2015). Mammograic image Analysis Society (MIAS) database v1.21.

[11] Rose, Chris, Daniele Turi, Alan Williams, Katy Wolstencroft, and Chris Taylor. "Web services for the DDSM and digital mammography research." In International Workshop on Digital Mammography, pp. 376-383. Springer, Berlin, Heidelberg, 2006.

[12] Moreira, Inês C., Igor Amaral, Inês Domingues, António Cardoso, Maria João Cardoso, and Jaime S. Cardoso. "Inbreast: toward a full-field digital mammographic database." Academic radiology 19, no. 2 (2012): 236-248.

[13] McCormack, V. A., & dos Santos Silva, I. (2006). Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiology and Prevention Biomarkers, 15(6), 1159-1169.

[14] Liberman, Laura, and Jennifer H. Menell. "Breast imaging reporting and data system (BI-RADS)." Radiologic Clinics 40.3 (2002): 409-430.

[15] Bovis, Keir, and Sameer Singh. "Classification of mammographic breast density using a combined classifier paradigm." In 4th international workshop on digital mammography, pp. 177-180. 2002.

[16] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." InComputer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 248-255. IEEE, 2009.

[17] M. Mu_stra, M. Grgi_c and K. Dela_c. A Novel Breast Tissue Density Classification Methodology. Breast Density Classi_cation Using Multiple Feature Selection. Automatika, Vol. 53 (4), pp. 362-372, 2012.

[18] W. He, E. Denton, K. Staord and R. Zwiggelaar. Mammographic Image Segmentation and Risk Classification Based on Mammographic Parenchymal Patterns and Geometric Moments. Biomed. Signal Processing and Control, vol.6(3), pp.321-329, 2011.

[19] Z. Chen, E. Denton and R. Zwiggelaar. Local feature based mamographic tissue pattern modelling and breast density classification. In the 4th Int. Conf. on Biomedical Engineering and Informatics, pp. 351-355, 2011.

[20] N. M. Parthalain, R. Jensen, Q. Shen and R Zwiggelaar. Fuzzy-rough approaches for mammographic risk analysis. Intelligent Data Analysis, Vol.14 (2), pp. 225-244, 2010.

[21] Fischer, E. A., J. Y. Lo, and M. K. Markey. "Bayesian networks of BI-RADS/spl trade/descriptors for breast lesion classification." In Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, vol. 2, pp. 3031-3034. IEEE, 2004.

[22] Ramos-Pollán, Raúl, Miguel Angel Guevara-López, Cesar Suárez-Ortega, Guillermo Díaz-Herrero, Jose Miguel Franco-Valiente, Manuel Rubio-del-Solar, Naimy González-de-Posada, Mario Augusto Pires Vaz, Joana Loureiro, and Isabel Ramos. "Discovering mammography-based machine learning classifiers for breast cancer diagnosis." Journal of medical systems 36, no. 4 (2012): 2259-2269.

[23] Wei, Chia-Hung, Yue Li, and Pai Jung Huang. "Mammogram retrieval through machine learning within BI-RADS standards."Journal of biomedical informatics 44, no. 4 (2011): 607-614.

[24] Arevalo, John, Fabio A. González, Raúl Ramos-Pollán, Jose L. Oliveira, and Miguel Angel Guevara Lopez. "Representation learning for mammography mass lesion classification with convolutional neural networks." Computer methods and programs in biomedicine 127 (2016): 248-257.

[25] Qiu, Yuchen, Shiju Yan, Maxine Tan, Samuel Cheng, Hong Liu, and Bin Zheng. "Computer-aided classification of mammographic masses using the deep learning technology: a preliminary study." In SPIE Medical Imaging, pp. 978520-978520. International Society for Optics and Photonics, 2016.

[26] Carneiro, Gustavo, Jacinto Nascimento, and Andrew P. Bradley. "Unregistered multiview mammogram analysis with pre-trained deep learning models." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 652-660. Springer, Cham, 2015.

[27] Shin, Hoo-Chang, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M. Summers. "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning." IEEE transactions on medical imaging 35, no. 5 (2016): 1285-1298.

[28] Bell, Anthony J., and Terrence J. Sejnowski. "The “independent components” of natural scenes are edge filters." Vision research37, no. 23 (1997): 3327-3338.

[29] Yamazaki, Yudai, Eiichi Takahashi, Masaya Iwata, Hirokazu Nosato, Ayumi Izumori, Takuji Iwase, Yumi Kokubu, and Hidenori Sakanashi. "Mammary gland segmentation using intensity correction for brightness attenuation in breast ultrasound image." In Imaging Systems and Techniques (IST), 2016 IEEE International Conference on, pp. 498-503. IEEE, 2016.

[30] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[31] He, Kaiming, 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, pp. 770-778. 2016.

[32] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521, no. 7553 (2015): 436-444.

[33] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets, Advances in neural information processing systems. 2014: 2672-2680.

[34] Bengio, Yoshua. "Learning deep architectures for AI." Foundations and trends® in Machine Learning 2.1 (2009): 1-127.

[35] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis. Proceedings of the 33 rd International Conference on Machine Learning, New York, NY, USA, 2016

[36] Neff, Thomas, Christian Payer, Darko Štern, and Martin Urschler. "Generative Adversarial Network based Synthesis for Supervised Medical Image Segmentation."