LATENT FEATURE REPRESENTATION WITH 3-D MULTI-VIEW DEEP CONVOLUTIONAL NEURAL NETWORK FOR BILATERAL ANALYSIS IN DIGITAL BREAST TOMOSYNTHESIS

Dae Hoe Kim, Seong Tae Kim and Yong Man Ro *

School of Electrical Engineering, KAIST, Daejeon, 305-701, Republic of Korea

ABSTRACT

In clinical studies of breast cancer, masses appear as asymmetric densities between the left and the right breasts, which show different breast tissue structures. For classifying breast masses, most researchers have developed hand-crafted bilateral features by extracting the asymmetric information in 2-D mammograms. In digital breast tomosynthesis (DBT), which has 3D volume data, effective bilateral features are needed to detect masses. In this paper, we propose latent bilateral feature representation with 3-D multi-view deep convolutional neural network (DCNN) in the DBT reconstructed volume. The proposed DCNN is designed to discover hidden or latent bilateral feature representation of masses in self-taught learning. Experimental results show that the proposed latent bilateral feature representation outperforms conventional hand-crafted features by achieving a high area under the receiver operating characteristic curve.

Index Terms**—** Digital breast tomosynthesis, deep learning, latent features, bilateral analysis, false positive reduction

1. INTRODUCTION

Breast cancer is the most common type of cancer in women [1]. Clinical reports have shown that an early detection of breast cancer significantly increases the survival rate [2]. Clinically, it is justified that the left and the right breast of the same patient tend to present a high degree of symmetry of the internal structures over broad areas, even if one breast may be larger than the other [3]. Therefore, the asymmetry between the left and right breasts of a given subject is an important sign used by radiologists to diagnose breast cancer and reduce false positives (FPs) [4]. Accordingly, in computer-aided detection (CAD) of masses in 2-D mammogram, many research efforts [5, 6] have been focused on the development of a bilateral analysis to improve mass detection sensitivity, and at the same time, reduce the number of FPs by utilizing the asymmetric characteristics of masses.

Recently, the importance of the digital breast tomosynthesis (DBT) has been increased, which has been designed to alleviate the tissue overlap problem occurring in Mammography [7]. The DBT provides a reconstructed volume that shows a cross-sectional slices (from 30 to 85 slices [8]) of the breast, which largely eliminates the tissue overlap problem [7]. However, for bilateral analysis in DBT, radiologists should compare the large number of slices of both breasts. This could induce a substantial increase in the workload and a possibility of overlooking subtle lesions. Thus, a computer-aided bilateral analysis in breast cancer screening on DBT is needed. Unlike the case of bilateral analysis in 2-D mammograms [5, 6], there are few research efforts devoted for bilateral analysis in DBT. Moreover, due to the increased data to be analyzed, effective features representation is needed in a bilateral analysis.

In this paper, we propose a latent bilateral feature representation learned from 3-D multi-view deep convolutional neural network (DCNN) for classifying masses and FPs. The idea of DCNN is to build hierarchical model which represents data at multiple levels of abstraction, which enables the model to discover accurate representations from data itself [9]. Thus, the DCNN is used to discover latent bilateral feature representation of masses, inherent in DBT volume. The rationale behind the proposed latent bilateral feature representation is based on the clinical fact that masses appear as an asymmetric volume of densities [3] which show different breast tissue structures between the left and right breasts. To learn asymmetric information of masses, a 3-D multi-view DCNN architecture is devised to learn image characteristics of mass volumes in main view and those of corresponding volumes in bilateral view. To the best of our knowledge, this work is the first attempt to devise an effective latent bilateral feature representation using deep learning for reconstructed volumes of the DBT. Experimental results show that the proposed latent bilateral mass feature representation achieve a higher level of classification performance in terms of the receiver operating characteristic (ROC) curves and the area under the ROC curve

* Corresponding author (ymro@ee.kaist.ac.kr)

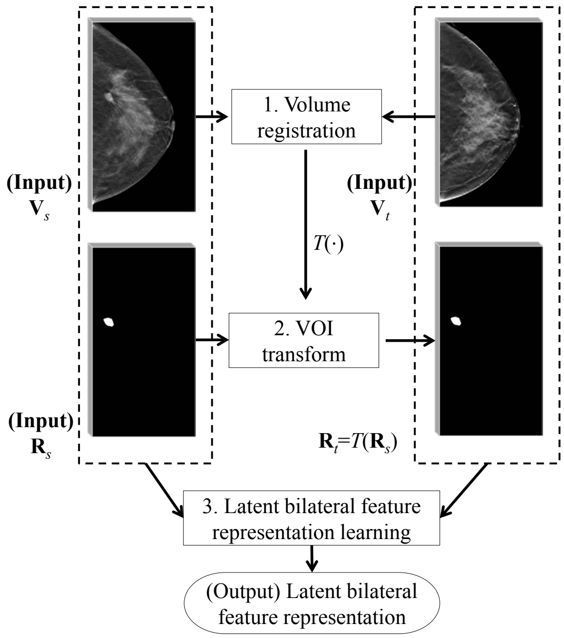

Figure 1. Overall process of the bilateral analysis using proposed latent bilateral feature representation with 3-D multi-view DCNN

(AUC) values [10] compared to the performance with hand-crafted features.

The rest of the paper is organized as follows. In section 2, the proposed bilateral analysis using latent bilateral feature representation with 3-D multi-view DCNN is explained. In section 3, experimental results are presented to verify the effectiveness of the proposed latent bilateral feature representation. The conclusions are drawn in section 4.

2. BILATERAL ANALYSIS USING LATENT BILATERAL FEATURE REPRESENTATION WITH 3-D MULTI-VIEW DCNN

Figure 1 shows the process of bilateral analysis using the proposed latent bilateral feature representation with 3-D multi-view DCNN. To learn the latent bilateral representation for a given VOI $\mathbf{R}_s$ in a source reconstructed volume $\mathbf{V}_s$, the corresponding VOI $\mathbf{R}_t$ in a target reconstructed volume $\mathbf{V}_t$ needs to be estimated. Let a geometric transformation function $T(\cdot)$ satisfy the following equation:

$$\mathbf{R}_t = T(\mathbf{R}_s). \tag{1}$$

The transform can be obtained as a result of the volume registration from $\mathbf{V}_s$ to $\mathbf{V}_t$. For this purpose, mutual information based volume registration with affine transformation in DBT reconstructed volume is adopted [11] which is known as one of the simple and effective registration methods in mammogram [12]. After the VOI transform, latent bilateral feature representation of masses is extracted from the $\mathbf{R}_s$ in the $\mathbf{V}_s$ and the $\mathbf{R}_t$ in the $\mathbf{V}_t$. Note that in the proposed method, transformed VOIs are used instead of the registered volume, to avoid the effect of registration error. The detail of the proposed latent bilateral feature representation using 3-D multi-view DCNN is described in the following subsection.

2.1. Latent bilateral feature representation of masses with 3-D multi-view DCNN

To learn latent bilateral features of masses in bilateral analysis, a 3-D multi-view DCNN architecture is devised. As shown in Figure 2, the architecture learns low-level feature representation of a mass VOI and the corresponding VOI separately. Then, both low-level feature representations are fused to learn high-level latent feature representation in the higher layer (F5).

The architecture of the DCNN is structured as a series of stages. The first four stages are composed of convolutional layers (C1, C1′, C3 and C3′) and subsampling layers (S2, S2′, S4 and S4′). For the given VOI $\mathbf{R}_s$ in the $\mathbf{V}_s$ and the corresponding VOI $\mathbf{R}_t$ in the $\mathbf{V}_t$, bounding boxes of both VOIs are cropped and resized to 32×32×25 size of volume. Then the pair of volumes is put in the proposed 3-D multi-view DCNN. When processing volumetric data, it is desirable to consider the intensity variation in depth direction. Thus, the convolutional layers (C1, C1′, C3 and C3′) adopt 3-D learnable filters as follows:

$$\mathbf{x}_j^{(l)} = f\left(\sum_{i=1}^{N^{(l)}} \mathbf{x}_i^{(l-1)} * \mathbf{W}_{ij}^{(l)} + \mathbf{b}_j^{(l)}\right), \tag{2}$$

where $\mathbf{x}_j^{(l)}$ is the j-th feature map in the l-th layer, $N^{(l)}$ is the number of feature map in the l-th layer, $*$ denotes a convolution operator, $\mathbf{W}_{ij}^{(l)}$ is the i-th 3-D filter of 5×5×5 size for the j-th feature map in the l-th layer and $\mathbf{b}_j^{(l)}$ is bias for the j-th feature map in the l-th layer. $f(\cdot)$ denotes an activation function. As the activation function, a rectified linear units (ReLU) [13] is employed. In the subsampling layers (S2, S2′, S4 and S4′), 2×2×2 maximum pooling is used with stride of 2 to make the network robust to local translation.

Essential features of mass VOIs and corresponding VOIs are captured through a series of convolutional layers and subsampling layers. The outputs of the second subsampling layers (S4 and S4′) are fused at a fully connected layer (F5) for the purpose of merging the feature responses of all feature maps. As a result, high-level latent bilateral features of the masses are encoded in higher layer (F5). In this paper, the output of fully connected layer is used as feature representation of the given VOI and the corresponding VOI of DBT volumes. The DCNN represents bilateral mass feature at multiple levels and discovers accurate representation from DBT data itself.

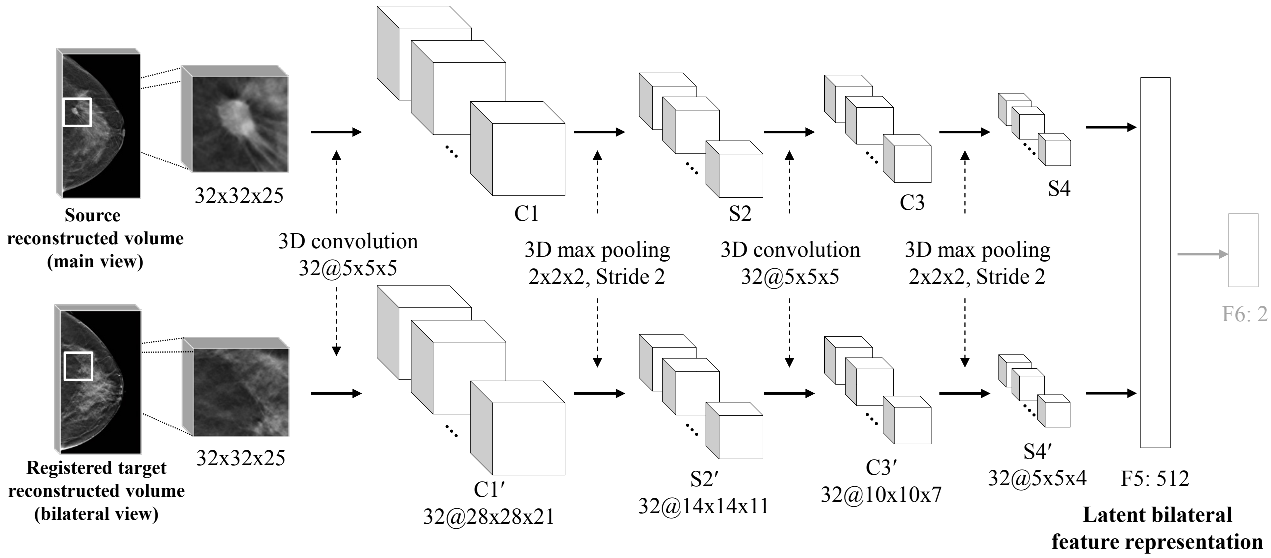

Figure 2. Proposed latent bilateral feature representation framework with 3-D multi-view DCNN

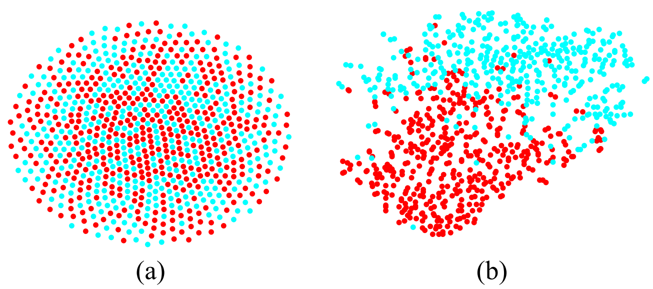

Figure 3. t-SNE feature visualization [14] of (a) original input data for the training set, (b) outputs of the fully-connected layer (F5) for the training data with the proposed DCNN. Red colored dots denote the mass samples and blue colored dots denote the normal breast tissues.

Figure 3 shows effectiveness of the proposed latent bilateral feature representation with t-distributed stochastic neighbor embedding (t-SNE) [14], which is a visualization technique for high-dimensional data. As shown in the figure, the original input data is spread over in the training data. The learned latent bilateral feature representations are relatively separated with respect to their class label. This indicates discriminative features are learned from raw data via the DCNN.

2.2. Training the 3-D multi-view DCNN for latent bilateral feature representation of masses

To train deep architecture, a large number of training data is required to avoid over-fitting. However, in medical imaging applications, it is difficult to collect a large number of true masses for the training. To overcome the problem, data augmentation is conducted by considering the characteristics of masses in DBT. In training set, mass VOIs are horizontally flipped. The original VOIs and flipped VOIs are rotated in [0°, 90°, 180° and 270°]. As a result, a total of 8 VOIs are generated from a single VOI.

In the training stage, the proposed 3-D multi-view DCNN minimizes classification errors in the last softmax layer (F6). In this paper, the classification error is calculated using a cross-entropy loss function as follows:

$$L = -\sum_{n=1}^{N_{tr}} \sum_{k=1}^{2} q_k^n \log p_k^n, \tag{3}$$

where N_{tr} denotes the number of training samples and p_k^n is the probability that the n-th training sample belongs to the k-th class. $q_k^n = 1$ if and only if the true class of the n-th training sample is k. The optimization is performed by minimizing the loss using stochastic gradient descent [15]. In this paper, the initial learning rate was set to 0.001, the momentum was set to 0.9 and the training epoch was set to 150. To avoid over-fitting, the fully-connected layer (F5) is constrained using drop out [15]. The probability of the drop out was set to 0.5 in this study.

3. EXPERIMENTS

A clinical data set of 160 reconstructed volumes was collected from Samsung Medical Center (with Hologic selenia dimensions® 3D system). The data set was acquired from craniocaudal (CC) and mediolateral oblique (MLO) views of both breasts of 40 patients. In 86 volumes, at least one biopsy-proven malignant mass was presented, while the remaining 74 volumes were interpreted as normal. The boundary of masses at the most conspicuous slice on each

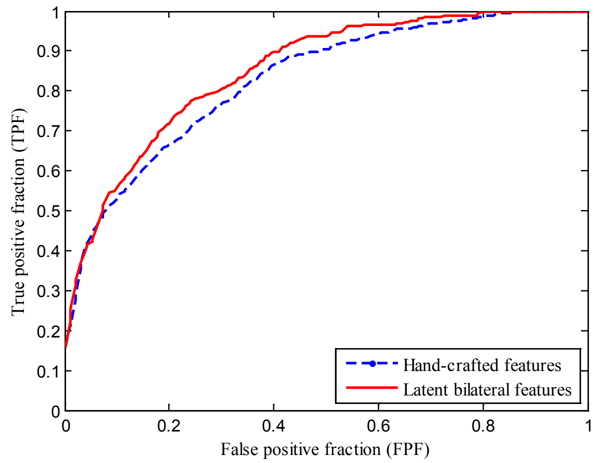

Figure 4. Comparisons of ROC curves of FP reduction using hand-crafted features and proposed latent bilateral features.

reconstructed volume was annotated by two experienced radiologists.

Table 1. Comparisons of averaged AUC between hand-crafted features and the proposed latent bilateral features

Used features	Averaged AUC	*p*-value
Hand-crafted features	0.826 ± 0.013	0.0102
Proposed latent bilateral features	0.847 ± 0.012	-

These annotations were used as a ground truth in this study. Each slice of reconstructed volume had the fixed pixel size of $0.1mm \times 0.1mm$ and a slice interval of $1mm$. For the computational efficiency, the reconstructed volumes were down-sampled to have the pixel size of $0.4mm$ x $0.4mm$ [16].

For a realistic assessment, VOIs of masses and FPs were automatically generated using a mass detection algorithm in the reconstructed volumes [8]. In the mass detection, the operating threshold was chosen which led to 78 mass VOIs and 479 normal tissue VOIs (i.e., FPs) (sensitivity of 90.7% at 2.99 FPs per breast volume). Note that, a VOI was considered as a mass if the centroid of the VOI fell within the boundary of the ground truth, or vice versa [16].

As for the classification, a support vector machine (SVM) with radial basis function was used to discriminate masses from FPs. The VOIs were randomly divided into two equal-sized groups of training and testing. In the training stage, 5-fold cross validation was applied for finding the optimal parameters of the SVM. The training and testing processes were repeated 50 times, and the averaged performances were presented as ROC curves and the AUC values [10].

For the comparison, a set of representative hand-crafted features on DBT reconstructed volume were used as 'hand-crafted features' in this paper: 354 region based multiresolution local binary pattern (LBP) features [17], 20 run length statistics (RLS) features [18], 52 spatial gray-level dependence (SGLD) features [19], 5 intensity features [20], and 433 bilateral features [11]. Note that for the hand-crafted features, a feature-level fusion [21] was adopted which is one of the representative methods for combining information residing in different feature spaces for pattern classification. As a result of the feature-level fusion, 864 dimensional feature vectors were generated.

Figure 4 and Table 1 shows the experimental results. As shown in the figure and the table, the classification performance of the proposed latent bilateral feature representation (AUC = 0.847) was higher than that of hand-crafted features (AUC = 0.826). Student's *t*-test was employed to measure the statistical significance of the improvement of AUC of the proposed latent feature representation. The proposed latent feature representation improved AUC of 0.021 (p=0.0102) compared to that of hand-crafted features. Note that the improvement of AUC is statistically significant if the *p*-value is less than 0.05 [8]. These results indicate that, in bilateral analysis, there are hidden or latent features for characterizing masses which conventional hand-crafted features could not capture. And the proposed latent bilateral feature representation can improve the overall classification performance in DBT bilateral analysis.

4. CONCLUSIONS

In this paper, we have proposed latent bilateral feature representation with 3-D multi-view DCNN in DBT reconstructed volume. The proposed DCNN is designed to discover hidden or latent bilateral feature representation of masses in self-taught learning. The experimental results showed that the proposed latent bilateral feature representation could improve the overall classification performance in terms of ROC and AUC in DBT.

5. REFERENCES

[1] D. M. Parkin, P. Pisani, and J. Ferlay, "Global cancer statistics," *CA: A Cancer Journal for Clinicians,* vol. 49, no. 1, pp. 33-64, Jan-Feb, 1999.

[2] T. W. Freer, and M. J. Ulissey, "Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center," *Radiology,* vol. 220, no. 3, pp. 781-6, Sep, 2001.

[3] D. B. Kopans, *Breast imaging*: Lippincott Williams & Wilkins, 2006.

[4] M. J. Homer, *Mammographic interpretation: A practical approach*, Boston, MA: McGraw-Hill, 1996.

[5] Y.-T. Wu, J. Wei, L. M. Hadjiiski, B. Sahiner, C. Zhou, J. Ge, J. Shi, Y. Zhang, and H.-P. Chan, "Bilateral analysis based false positive reduction for computer-aided mass detection," *Medical physics,* vol. 34, no. 8, pp. 3334-3344, 2007.

[6] F. F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images," *Medical Physics,* vol. 18, no. 5, pp. 955-963, 1991.

[7] J. A. Baker, and J. Y. Lo, "Breast tomosynthesis: state-of-the-art and review of the literature," *Academic Radiology,* vol. 18, no. 10, pp. 1298-310, Oct, 2011.

[8] S. T. Kim, D. H. Kim, and Y. M. Ro, "Breast mass detection using slice conspicuity in 3D reconstructed digital breast volumes," *Physics in Medicine and Biology,* vol. 59, no. 17, pp. 5003-5023, 2014.

[9] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends® in Machine Learning,* vol. 2, no. 1, pp. 1-127, 2009.

[10] C. E. Metz, "Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems," *Journal of the American College of Radiology,* vol. 3, no. 6, pp. 413-422, 2006.

[11] D. H. Kim, S. T. Kim, and Y. M. Ro, "Feature extraction from bilateral dissimilarity in digital breast tomosynthesis reconstructed volume (Accepted)," *in Image Processing (ICIP), IEEE International Conference on*, 2015.

[12] P. R. Snoeren, and N. Karssemeijer, "Gray-scale and geometric registration of full-field digital and film-screen mammograms," *Medical image analysis,* vol. 11, no. 2, pp. 146-156, 2007.

[13] V. Nair, and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *in Proc. 27th Int'l Conf. on Machine Learning.* pp. 807-814, 2010.

[14] L. Van der Maaten, and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.,* vol. 9, no. 2579-2605, pp. 85, 2008.

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

[16] H. P. Chan, J. Wei, Y. Zhang, M. A. Helvie, R. H. Moore, B. Sahiner, L. Hadjiiski, and D. B. Kopans, "Computer-aided detection of masses in digital tomosynthesis mammography: comparison of three approaches," *Medical physics,* vol. 35, no. 9, pp. 4087-95, Sep, 2008.

[17] J. Y. Choi, and Y. M. Ro, "Multiresolution local binary pattern texture analysis combined with variable selection for application to false-positive reduction in computer-aided detection of breast masses on mammograms," *Physics in Medicine and Biology,* vol. 57, no. 21, pp. 7029-7052, 2012.

[18] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer graphics and image processing,* vol. 4, no. 2, pp. 172-179, 1975.

[19] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on,* vol. SMC-3, no. 6, pp. 610-621, 1973.

[20] H. Cheng, X. Shi, R. Min, L. Hu, X. Cai, and H. Du, "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognition,* vol. 39, no. 4, pp. 646-668, 2006.

[21] U. G. Mangai, S. Samanta, S. Das, and P. R. Chowdhury, "A survey of decision fusion and feature fusion strategies for pattern classification," *IETE Technical Review,* vol. 27, no. 4, pp. 293-307, 2010.