Automatic Dual-View Mass Detection in Full-Field Digital Mammograms

Guy Amit, Sharbell Hashoul, Pavel Kisilev, Boaz Ophir,
Eugene Walach, and Aviad Zlotnick

IBM Research - Haifa, Israel

Abstract. Mammography is the first-line modality for screening and diagnosis of breast cancer. Following the common practice of radiologists to examine two mammography views, we propose a fully automated dual-view analysis framework for breast mass detection in mammograms. The framework combines unsupervised segmentation and random-forest classification to detect and rank candidate masses in cranial-caudal (CC) and mediolateral-oblique (MLO) views. Subsequently, it estimates correspondences between pairs of candidates in the two views. The performance of the method was evaluated using a publicly available full-field digital mammography database (INbreast). Dual-view analysis provided area under the ROC curve of 0.94, with detection sensitivity of 87% at specificity of 90%, which significantly improved single-view performance (72% sensitivity at 90% specificity, 78% specificity at 87% sensitivity, $P<0.05$). One-to-one mapping of candidate masses from two views facilitated correct estimation of the breast quadrant in 77% of the cases. The proposed method may assist radiologists to efficiently identify and classify breast masses.

Keywords: Digital Mammography, Automatic Mass Detection, Dual-View, Machine Learning.

1 Introduction

Mammography is the most common imaging modality for breast cancer screening, with over 38 million tests performed annually in the US. The majority of certified breast imaging facilities in the US already use full-field digital mammography (FFDM) [1]. When interpreting a mammogram, radiologists typically examine images obtained from two anatomical projections: the cranial-caudal (CC) view and the mediolateral-oblique (MLO) view. A breast mass is a lesion seen in both views. However, since the two projected images are acquired by applying different compression on the non-rigid breast tissue, deriving the dual-view correspondence between potential masses is a non-trivial task.

Computer-aided detection of breast masses in mammography has been studied extensively. Earlier work focused on single-view analysis using either supervised, model-based methods or unsupervised techniques, including region-based segmentation, boundary detection, and pixel clustering (see [2] for a thorough review). Previous

N. Navab et al. (Eds.): MICCAI 2015, Part II, LNCS 9350, pp. 44–52, 2015.
DOI: 10.1007/978-3-319-24571-3_6

work on multi-view mass detection in mammograms introduced various correspondence features including distance to the nipple, region- and histogram correlations, and texture gradients [3, 4]. These features were used either to restrict the search region in the second view [5, 6] or to jointly classify pairs of masses, using machine learning techniques such as linear discriminant analysis [3, 7], artificial neural networks [4] or Bayesian networks [8]. The mammograms used in the abovementioned work were either proprietary image datasets, used for a particular research, or publicly available datasets such as DDSM and MIAS, which consist of digitized film images. As FFDM becomes the standard clinical technology, automated analysis methods should be adapted [9] and evaluated on FFDM benchmarks. INbreast (courtesy of Breast Research Group, INESC Porto, Portugal) is a new publicly-available full-field digital mammography database, which includes detailed annotations of masses [10].

The current work introduces a new fully automated dual-view analysis framework for breast mass detection in FFDM images. Our work entails three major contributions: (1) A hybrid unsupervised-supervised approach for segmentation and ranking of candidate masses in single-view images; (2) A learning-based method for estimating the correspondence between candidate masses in two views; (3) An evaluation of the proposed automatic detection methods on the INbreast database.

2 Methods

2.1 Single View Detection

The single view mass detection pipeline [11] consists of image preprocessing, detection of candidate seeds, generation of candidate contours, feature extraction, and candidate ranking (Fig 1a).

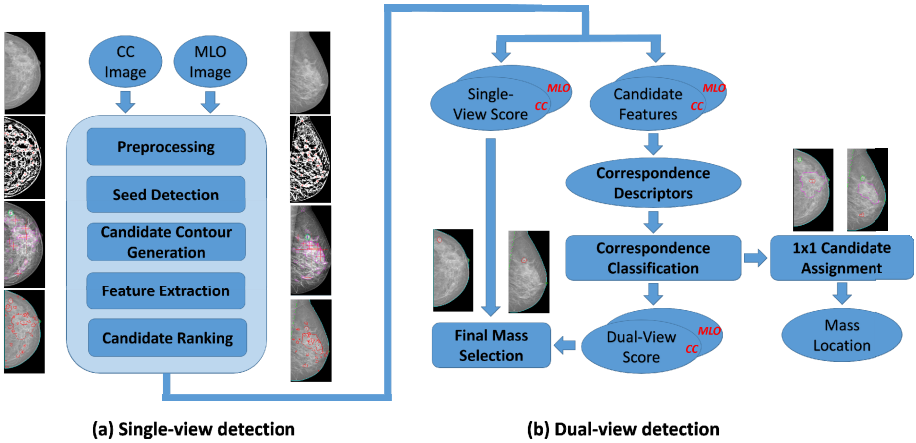

Fig. 1. Single-view analysis framework (a) detects and ranks mass candidates separately in CC and MLO images. The dual-view correspondence between pairs of candidates is estimated and a combined dual-view score is used to select true masses and reject false-positives (b). Dual-view candidate assignment is also used to estimate mass location.

Images were first preprocessed by separating the breast from the background and identifying the pectoral muscles and the nipple. Seed detection was based on applying semantic thresholding [12], followed by a distance transform. Semantic thresholding classifies image pixels as locally dark or locally bright, and finds a global binarization threshold that balances the number of locally dark pixels that are brighter than the threshold and locally bright pixels that are darker than the threshold. The image is partitioned into 20x20 tiles and a threshold is computed for each tile. The resulting 20x20 threshold matrix is smoothed by a Gaussian filter, interpolated to the image dimensions and used as a per-pixel threshold for binarization (Fig. 2b). A distance transform is then applied to the binary image (Fig 2c) and large connected components are selected as candidate seeds, using experimentally-set thresholds (Fig. 2d). The selected seeds are sorted by the maximal value of the distance transform in the connected component. This value is used as an initial unsupervised detection score, denoted S_i^0. To account for large lesions, the process is repeated with a coarser image portioning (5x5 tiles), and the largest connected component is added to the selected seeds.

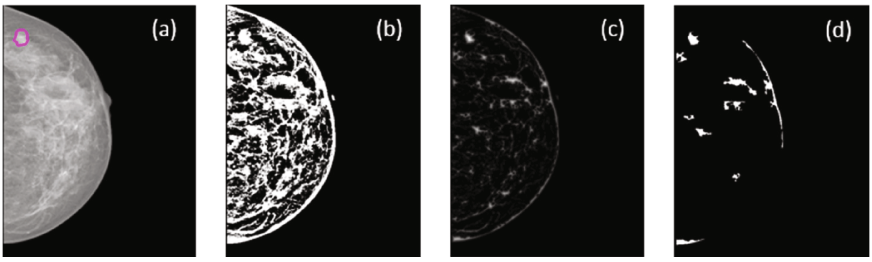

Fig. 2. Seed detection. A mammogram (a) is processed by semantic thresholding to produce a binary image (b). A distance transform is applied (c), and large connected components are selected (d). The contour annotation in (a) indicates a ground-truth mass.

The contours of mass candidates were generated by finding an optimal path in a polar representation of the image around the seeds, similar to [13] and [14]. The maximal value of the seed's distance transform was used as an initial estimate for the radius of the polar image.

For each ground-truth and candidate mass contour, we extract a multitude of features, including: (1) Shape features: area, aspect ratio, curvature along the boundaries, orientation, and eccentricity of a fitted ellipse; (2) Intensity features: intensity statistics, normalized intensity histograms inside and outside the contour; (3) Texture features: local entropy of gray-level values, calculated at different scales [15].

Mass ranking was done with a random forest (RF) classifier [16] using a leave-one-patient-out cross validation scheme. For each patient, a model was trained using a balanced set of true- and false-detected mass contours from all other patients. True-detection observations were the ground-truth contours, m_{gt}, as well as detected candidate masses m_i with large ground-truth overlap $dice(m_i, m_{gt}) \geq \delta$, where $dice(A,B) = 2|A \cap B|/(|A| + |B|)$ and $\delta = 0.5$. False-detection observations were candidate masses without any ground-truth overlap$(dice(m_i, m_{gt}) = 0)$. For each

test candidate contour, a class label $l(m_i) \in \{0,1\}$ was predicted, classifying the candidate as a true or false mass. The candidate's *single-view score* $S_i^1 \in [-1,1]$ was defined by $S_i^1 = \mathcal{P}(l(m_i) = 1) - \mathcal{P}(l(m_i) = 0)$, where $\mathcal{P}(l(m_i) = \mathcal{L})$ is calculated as the fraction of trees in the random-forest model that voted for class label $\mathcal{L}$.

2.2 Dual View Detection

The analysis framework for detecting corresponding mass candidates in two views is illustrated in Figure 1b. Given a pair of CC and MLO mammograms $I = (I^{CC}, I^{MLO})$, each having a set of candidate masses $M^{CC} = \{m_1^{CC}, \dots, m_N^{CC}\}$, $M^{MLO} = \{m_1^{MLO}, \dots, m_N^{MLO}\}$, the analysis aims to find a relation from M^{CC} to M^{MLO} that contains all the corresponding candidate pairs from the two views. A *correspondence descriptor* D_{ij} is a dual-view feature vector encoding the matching between single-view features of m_i^{CC} and m_j^{MLO}. D_{ij} consists of the following features: (1) Location features: ratio of the distances between the centroid of the candidate mass and the nipple, ratio of the distances between the centroid of the candidate mass and a line tangential to the nipple and parallel to the pectoral muscle; (2) Shape features: ratios between single-view shape features, including area, perimeter, eccentricity, ellipse axis-length, solidity; (3) Image features: contrast difference, differences of intensity statistics, histogram distance (Earth Mover's Distance) and histogram flow [17].

The correspondence between candidate masses m_i^{CC} and m_j^{MLO} is assessed by a *correspondence score* $C_{ij} \in [-1,1]$. To estimate C_{ij}, a classifier is trained and used to predict a label $l(D_{ij}) \in \{0,1\}$ for each correspondence descriptor D_{ij}. The *correspondence score* is defined by: $C_{ij} = \mathcal{P}\big(l\big(D_{ij}\big) = 1\big) - \mathcal{P}\big(l\big(D_{ij}\big) = 0\big)$.

The classifier's training data is composed of correspondence descriptors of matching ('*positive*') and non-matching ('*negative*') pairs of masses. For an image pair *I*, D_{gt}^I is the correspondence descriptor of the ground-truth mass contours $(m_{gt}^{CC}, m_{gt}^{MLO})$. The set of positive observations in our training set consists of ground-truth descriptors, as well as the pairs of candidate contours with sufficiently-large overlap with the ground-truth:

$$Observ_+^I = \{D_{gt}^I\} \cup \{D_{ij}^I \big| dice\big(m_i^{CC}, m_{gt}^{CC}\big) \geq \delta, dice\big(m_j^{MLO}, m_{gt}^{MLO}\big) \geq \delta\}$$

Likewise, the set of negative observations consists of correspondence descriptors of candidate pairs, where the contour in one view overlaps with the ground-truth, while the contour in the other view is a false detection, which does not overlap with the ground-truth:

$$\begin{aligned} Observ_-^I = &\{D_{ij}^I \big| dice\big(m_i^{CC}, m_{gt}^{CC}\big) \geq \delta, dice\big(m_j^{MLO}, m_{gt}^{MLO}\big) = 0\} \\ &\cup \{D_{ij}^I \big| dice\big(m_i^{CC}, m_{gt}^{CC}\big) = 0, dice\big(m_j^{MLO}, m_{gt}^{MLO}\big) \geq \delta\} \end{aligned}$$

Similar to mass classification, the correspondence descriptors of each image pair were classified by a random-forest classifier using leave-one-patient-out cross validation.

The likelihood of a test candidate having a corresponding mass candidate in the other view was estimated by a *dual-view score* $S_i^2 = \max_j\{C_{ij}\}$.

To estimate the spatial location of the candidate masses, a one-to-one correspondence between pairs of candidates was derived by solving the linear assignment problem, minimizing $\sum_i \sum_j (1 - C_{ij}) x_{ij}$, where $x_{ij} = \begin{cases} 1, & if\ m_i^{CC}\ is\ assigned\ to\ m_j^{MLO} \\ 0, & otherwise \end{cases}$

2.3 Evaluation Experiments

Our analysis included 90 images from 43 patients. All selected images had a single annotated mass. In order to simplify the interpretation of the results, images with multiple mass annotations or cases with missing annotation in one of the views were excluded (18 images). For each image, unsupervised seed detection was applied, and candidate mass contours were generated from the 30 first seeds, ordered by the detection score, S_i^0. A seed that resided within the ground-truth contour was considered a true detection. The candidate contours were then ranked by computing the single-view score S_i^1, using leave-one-patient-out cross validation. The quality of the candidate contours was assessed by their Dice-index overlap with the ground-truth contours. The single-view detection performance was measured by the area under the receiver operating characteristic curve (AUROC) and the detection sensitivity at a false-positive rate (FPR) of 10%. Dual-view analysis was carried out using the first 10 mass candidates from each view. For each pair of CC-MLO images, the analysis consisted of classification of correspondence descriptors, calculation of dual-view scores S_i^2 and breast quadrant estimation, following one-to-one candidate assignment. The AUROC of the combined dual-view score $S_i^1 + S_i^2$ was compared to the single-view score, and the differences in detection sensitivity and specificity were assessed using McNemar's chi-square test.

3 Results

Sorting the detected seeds by the unsupervised score S_i^0, a true detection was found within the first 30 seeds, 10 seeds and 1 seed in 96%, 79% and 37% of the images, respectively. The average Dice-index overlap between the best automatically detected candidate mass contour and the ground-truth mass contour was 0.8±0.2. This overlap was > 0.5 in 96% of the images. Candidate mass classification with S_i^0 achieved an AUROC of 0.71, with sensitivity of 36% at specificity of 90% (Fig 3).

Following learning-based candidate ranking, the 10 highest-ranked candidates included a true detection (Dice > 0.5) in 89% of the images. In 70% of the images the first-ranked candidate was a true detection, which corresponds to an average FPR of 0.3 per image. With 2.2 false-positives per image, the detection sensitivity was 83%. The AUROC of candidate classification by the single-view score S_i^1 was 0.92 (Fig. 3), with detection sensitivity of 72% at a specificity of 90%.

Classification of the correspondence descriptors into true- and false pairwise matches provided an AUROC of 0.96, with optimal sensitivity and specificity of 89%

and 96%, respectively. Following the one-to-one candidate assignment, correct pairs of true CC-MLO mass candidates were found in 67% of the pairs, and the correct breast quadrant was estimated in 77% of the cases.

The combined dual-view score $S_i^1 + S_i^2$ improved the candidate classification compared to the single-view score (Fig 3), with AUROC of 0.94 and detection sensitivity of 87% vs. 72%, at specificity of 90% (P<0.05). At the same detection sensitivity of 87%, the specificity of the single-view score was lower at 78% (P<0.05).

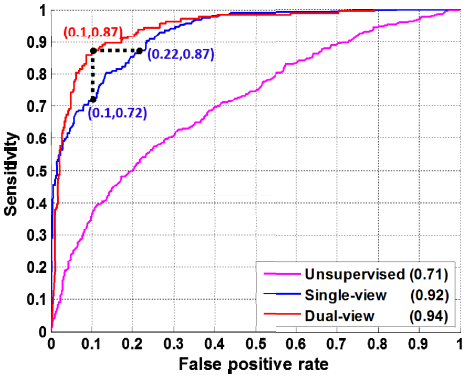

Fig. 3. ROC curves and AUROC of unsupervised- single-view- and combined dual-view detection scores. The marked points indicate the differences in sensitivity and specificity between single- and dual-view analyses.

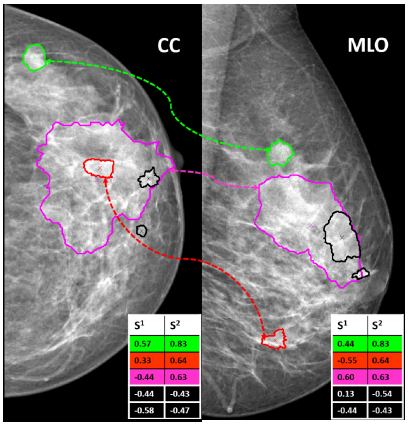

Fig. 4. CC-MLO candidate correspondence example. S^1 and S^2 are the single- and dual-view scores for each candidate. A combined score is used to reject unpaired candidates (black) and select the best pair (green).

Figure 4 shows a typical example of dual-view correspondence between detected candidates. In this example, three pairs of candidates were linked with high dual-view scores, while two additional candidates in each view were rejected as false detections. The combination of S_i^1 and S_i^2 scores enabled correct selection of the true mass candidate (green) in both views, although its MLO single-view score was not the maximal.

4 Discussion

Oliver et al. [2] evaluated seven different single-view detection algorithms, using two mammography databases. The reported AUROC ranged from 0.56 to 0.79. None of the evaluated algorithms provided an overall best performance, and there was a trend of decreased performance when using FFDM images, compared to digitized mammograms. Wei et al. [18] analyzed a proprietary FFDM dataset and reported sensitivity of 70% at a FPR of 0.72 per image, using morphological and gray-level texture features and a combination of rule-based and linear discriminant analysis (LDA)

classifiers. Paquerault et al. [7] reported single-view sensitivity of 62%, with 1 false-positive per image, while their two-view detection scheme increased the sensitivity to 73%. Van Engeland et al. [3] used LDA to classify correct and incorrect pairs of suspicious mass regions. Using features such as difference of mass-nipple distances, contrast difference and histogram correlation, a correct pairing of true positive regions was found in 82% of the cases. Yuan et al. [4] applied a Bayesian artificial neural network to estimate the probability of correspondence between lesions from different views, reporting AUROC of 0.87. Velicova et al. [8] proposed a fixed Bayesian network framework for modeling multi-view dependences, which increased the AUROC by 0.01 to 0.05. Recently, Kozegar et al. [19] assessed their mass detection method using INbreast mammograms. Their proposed ensemble of classifiers achieved sensitivity of 87% at FPR of 3.67 per image. At a lower FPR of 0.5 per image, the sensitivity decreased to 32%. Our single-view analysis framework, based on a similar approach achieved 87% sensitivity at a lower FPR of 3.13 per image, and maintained a fair detection sensitivity of 70% at a low FPR of 0.3 per image. Fusing the dual-view correspondence score with the single-view ranking score boosted the detection performance, reducing either the false-negative or the false-positive rate by more than half, compared to the single-view performance (Fig. 3). The pairwise correspondence classification was highly accurate, with AUROC of 0.96. These results strengthen the potential of the automated dual-view analysis. Our proposed approach also provided an estimation of the three-dimensional spatial location of the mass, based on deriving one-to-one correspondence between candidate masses. As breast quadrants have different likelihoods for incidence of cancer, accurate estimation of the quadrant may be useful for further semantic classification of the detected masses into benign or malignant lesions.

While training a random forest classifier, the relative importance of each feature can be measured by randomly perturbing every variable and computing the average difference in accuracy. In our experiments, the most prominent features for two-view correspondence classification were the ratios of mass-nipple distances, mass areas, ellipse axis lengths, normalized distances to the center of the image, and earth mover's distance between the local histograms. The proposed analysis framework is highly scalable in terms of the feature set used for single- and dual-view detection, enabling experimentation with new informative features, as well as with feature selection strategies.

The use of INbreast FFDM database to assess the framework's performance takes advantage of the detailed ground-truth mass annotations, which allows the definition of fine shape correspondences between annotated contours. Moreover, the major benefit of using INbreast lies in its availability to the research community, allowing the use of standard benchmarking of mammogram analysis algorithms. A current limitation of this dataset is its relatively-small number of mass-containing images. This limitation was addressed by evaluating the classifiers using leave-one-out cross validation. However, further validation of the method on additional datasets is needed. Another limitation of this work is the restriction of analysis to mammograms with a single annotated mass. A natural extension would be to evaluate the method on multiple-mass images, as well as images with different numbers of masses in each view.

The proposed dual-view analysis framework could assist radiologists in decreasing their workload, by automatically indicating suspicious regions that require attention. Such fusion of human knowledge and computer algorithms bears a true promise for future cognitive systems in diagnostic radiology.

References

1. Mammography Quality Standards Act's National Statistics. http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/
2. Oliver, A., Freixenet, J., Martí, J., Pérez, E., Pont, J., Denton, E.R.E., Zwiggelaar, R.: A review of automatic mass detection and segmentation in mammographic images. Med. Image Anal. 14, 87–110 (2010)
3. Van Engeland, S., Timp, S., Karssemeijer, N.: Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views. Med. Phys. 33, 3203–3212 (2006)
4. Yuan, Y., Giger, M.L., Li, H., Sennett, C.: Correlative feature analysis on FFDM. Med. Phys. 35, 5490–5500 (2008)
5. Zheng, B., Leader, J.K., Abrams, G.S., Lu, A.H., Wallace, L.P., Maitz, G.S., Gur, D.: Multiview-based computer-aided detection scheme for breast masses. Med. Phys. 33, 3135–3143 (2006)
6. Wiemker, R., Kutra, D., Heese, H., Buelow, T.: Identification of corresponding lesions in multiple mammographic views using star-shaped iso-contours. In: Aylward, S., Hadjiiski, L.M. (eds.) SPIE Medical Imaging, p. 90351A. International Society for Optics and Photonics (2014)
7. Paquerault, S., Petrick, N., Chan, H.-P., Sahiner, B., Helvie, M.A.: Improvement of computerized mass detection on mammograms: fusion of two-view information. Med. Phys. 29, 238–247 (2002)
8. Velikova, M., Samulski, M., Lucas, P.J.F., Karssemeijer, N.: Improved mammographic CAD performance using multi-view information: a Bayesian network framework. Phys. Med. Biol. 54, 1131–1147 (2009)
9. Li, H., Giger, M.L., Yuan, Y., Chen, W., Horsch, K., Lan, L., Jamieson, A.R., Sennett, C.A., Jansen, S.A.: Evaluation of computer-aided diagnosis on a large clinical full-field digital mammographic dataset. Acad. Radiol. 15, 1437–1445 (2008)
10. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INbreast: Toward a Full-field Digital Mammographic Database. Acad. Radiol. 19, 236–248 (2012)
11. Zlotnick, A., Ophir, B., Kisilev, P.: Hybrid Unsupervised-Supervised Lesion Detection in Mammograms. SPIE Medical Imaging (2015)
12. Zlotnick, A., Lozinskii, E.: Semantic thresholding. Pattern Recognit. Lett. 5, 321–328 (1987)
13. Timp, S., Karssemeijer, N.: A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. Med. Phys. 31, 958–971 (2004)
14. Rojas Domínguez, A., Nandi, A.K.: Improved dynamic-programming-based algorithms for segmentation of masses in mammograms. Med. Phys. 34, 4256–4269 (2007)
15. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn (2006)

16. Breiman, L.: Random Forests. Mach. Learn. 45, 5–32 (2001)
17. Kisilev, P., Freedman, D., Wallach, E., Tzadok, A., Naveh, Y.: DFlow and DField: New features for capturing object and image relationships. In: 21st International Conference on Pattern Recognition (ICPR), pp. 3590–3593. IEEE (2012)
18. Wei, J., Sahiner, B., Hadjiiski, L.M., Chan, H.-P., Petrick, N., Helvie, M.A., Roubidoux, M.A., Ge, J., Zhou, C.: Computer-aided detection of breast masses on full field digital mammograms. Med. Phys. 32, 2827–2838 (2005)
19. Kozegar, E., Soryani, M., Minaei, B., Domingues, I.: Assessment of a novel mass detection algorithm in mammograms. J. Cancer Res. Ther. 9, 592–600 (2013)