AUTOMATIC DETECTION OF REGION OF INTERESTS IN MAMMOGRAPHIC IMAGES

Erkang Cheng[1] Haibin Ling[1] Predrag R. Bakic[3]
Andrew D.A. Maidment[3] Vasileios Megalooikonomou[2]

[1] Center for Information Science and Technology, Temple University
[2] Data Engineering Laboratory (DenLab), Temple University
[3] Department of Radiology, University of Pennsylvania

ABSTRACT

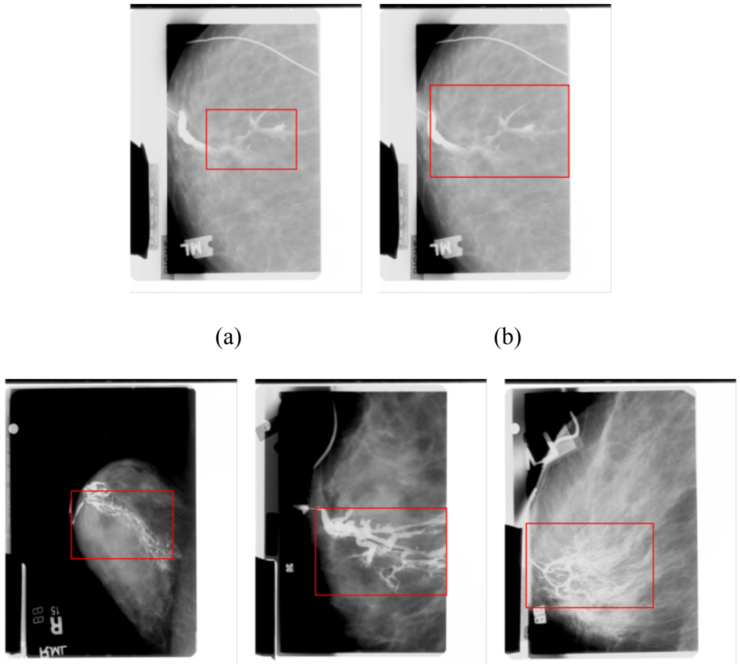
This work is a part of our ongoing study aimed at comparing the topology of anatomical branching structures with the underlying image texture. Detection of regions of interest (ROIs) in clinical breast images serves as the first step in development of an automated system for image analysis and breast cancer diagnosis. In this paper, we have investigated machine learning approaches for the task of identifying ROIs with visible breast ductal trees in a given galactographic image. Specifically, we have developed boosting based framework using the AdaBoost algorithm in combination with Haar wavelet features for the ROI detection. Twenty-eight clinical galactograms with expert annotated ROIs were used for training. Positive samples were generated by resampling near the annotated ROIs, and negative samples were generated randomly by image decomposition. Each detected ROI candidate was given a confidences core. Candidate ROIs with spatial overlap were merged and their confidence scores combined. We have compared three strategies for elimination of false positives. The strategies differed in their approach to combining confidence scores by summation, averaging, or selecting the maximum score.. The strategies were compared based upon the spatial overlap with annotated ROIs. Using a 4-fold cross-validation with the annotated clinical galactographic images, the summation strategy showed the best performance with 75% detection rate. When combining the top two candidates, the selection of maximum score showed the best performance with 96% detection rate

Keywords: Mammography, Galactography, Machine learning techniques

1. INTRODUCTION

Mammographic image analysis is an important tool for clinical breast cancer diagnosis and therapy validation. Despite recent advances in computer-aided analysis, initial identification of a region of interest (ROI) from a clinical image usually requires a manual or semi-manual intervention from an expert. Such identified region is used for further task-dependent processing. The manual ROI localization by an expert, although accurate, represents an obstacle to a fully automatic analysis. The position of such an ROI in a clinic study may be very flexible, as long as the diagnostically important structures are included within the ROI. For example, the first two images in Fig. 1 show different ROI annotations from two experts on the same galactogram (*i.e.*, a mammographic projection of the breast with contrast agent injected to enhance a portion of the ductal network). This work is a part of our ongoing study aimed at comparing the topology of anatomical branching structures with the underlying image texture. Detection of ROIs in clinical mammograms and galactograms of the same breast would serve as the first step in an automated system for analysis of topological and textural properties.

Medical Imaging 2011: Image Processing, edited by Benoit M. Dawant, David R. Haynor,
Proc. of SPIE Vol. 7962, 79623J · © 2011 SPIE · CCC code: 1605-7422/11/$18 · doi: 10.1117/12.877667



(a) (b)

Figure 1. Example mammographic images
(Note that the image (a), (b) are the same but have different annotations.)

The challenge of automatic ROI detection, in mammographic images mainly lies in large variations of texture and shape of anatomical structures depicted in the ROIs. Such variations comes from large deformability of the breast tissue,, systematic effects during the imaging process, and potential pathologic changes in the breast anatomy. In addition, as shown in Fig. 1, the expected position of anatomical structures visualized by an ROI varies from image to image, which makes atlas based approach inappropriate. Motivated by recent advance in object detection in computer vision and medical imaging, we propose applying machine learning approaches for automated ROI detection. These approaches are capable of capturing the variation in data and therefore become robust to the above mentioned challenges.

Galactography is a clinical modality for visualization of breast ductal network [12], and is indicated in cases of nipple discharge with no palpable or mammographically visible lesions. Analysis of galactograms has been used in simulation of breast ductal network, as a part of developing an anthropomorphic software breast phantom [13]. Galactograms have been analyzed also for comparison between the ductal topology and the corresponding mammographic texture [14]. Similarly, methods for ROI based calculation of the pixel intensity histogram [6] were tested for automated mammogram analysis.

In our previous work, we have used local hybrid features to detect nodes in a graph representing ductal network in galactograms [10, 11]. Combining a classifier trained by a support vector machine (SVM) or AdaBoost, the local hybrid features were utilized to predict a probabilistic map of graph nodes.

Object detection from visual input has achieved great progress in the past decade and has been successfully applied to tasks such as face and vehicle detection [4] using machine learning approaches. Similar approaches have recently been applied to anatomical structure localization tasks as well. A pixel based method using edge and intensity as features was examined in [8]. The classifier is just based on threshold learning from training data. In another application [9] a neural network was used to learn a classifier based on geometry intensity.

We have used the AdaBoost algorithm [5] in combination with Haar wavelets for the task of automated ROI detection. Combining AdaBoost with Haar wavelets has been shown to be very effective for face detection in [4]. While the method works well for detection of general object such as faces, several issues needed to be addressed in our task. First, the database of clinical galactograms used in our task had a limited number of training samples. We solved this problem by sampling more ROIs that near an annotated one; we also decomposed original images to generating more negative samples. Second, in our application, we assumed that there always exists only one ROI in each input image. As a result, we investigate several strategies to remove false positives. The proposed method has been tested on a dataset of 28 clinical galactographic images.

The rest of the paper is organized as follows: in Section 2 we introduce the proposed learning based method for ROI detection. Then, we present the experimental evaluation in Section 3. Finally, Section 4 shows the experiments results and Section 5 concludes the paper.

2. METHODOLOGY

2.1 AdaBoost for ROI Detection

AdaBoost is an efficient ensemble learning method that is first proposed by Freund and Schapire[5]. The method has been widely used in many machine problems. Given a set of training samples, the basic idea is to build a strong classifier as a linear combination of a set of weak classifiers.

Specifically, let $x \in \Re^d$ be a d dimensional input feature vector. Here, a feature vector is extracted to represent a training instance. Actually, a weak classifier could be obtained if only one special dimension of the feature vector is considered. AdaBoost could achieve a classifier with better performance which combines a set of weak classifiers. In the beginning of the training stage, each example is initialized with an equalized weight. A distribution of these weights is updated during the training process. AdaBoost selects weak classifier repeatedly by update the distribution of these weights. The distribution of weights indicates the importance of examples in the data set. On each training round, the weight of each incorrectly classified example are increased, the weight of each correctly classified example are decreased. Therefore, the new weak classifier in next round will focus more effort on those incorrect classified instances. AdaBoost could stop by setting loop rounds or when the classifier satisfies certain classification rate. Generally, the a classifier $h(x)$: $\Re d \rightarrow \{-1,1\}$

$$h(x) = \mathrm{sign}(\textstyle\sum_{i=1..n} c_i\, h_i(x)) \tag{1}$$

where $h_1(x)$, …, $h_n(x)$ are the n weak classifiers got in the training stage, c_i is the weight of each weak classifier.

In our task, we assumed that each image has one and only one ROI, therefore we can use a confidence score for each detection. In other word, instead of a label (1 or -1) given in (1), the detection score is used for filter out false positive. From (1), we define the following score function $f(x)$:$\Re d \rightarrow \Re$

$$f(x) = \textstyle\sum_{i=1..n} c_i\,(x)\ . \tag{2}$$

The confidence score is combined in the detection stage to search the optimal candidate.

In our work, we use Haar-like features with AdaBoost. Haar-like features (Haar features for short) have been successfully used in the first real-time face detector [4]. Taking benefit of the integral images, a Haar feature can be calculated efficiently in constant time. The computation of Haar features is through the Haar-like filters as shown in Fig. 2. There are seven Haar-like filters with different shapes. In this figure, the sum of the pixels lying inside the white rectangles is subtracted from the sum of the pixels lying within the grey rectangles.

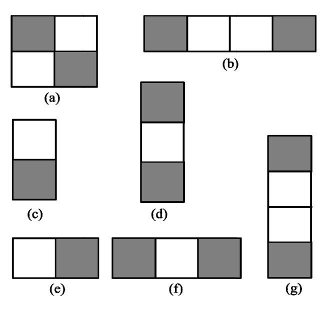


Figure. 2: Haar-like filters

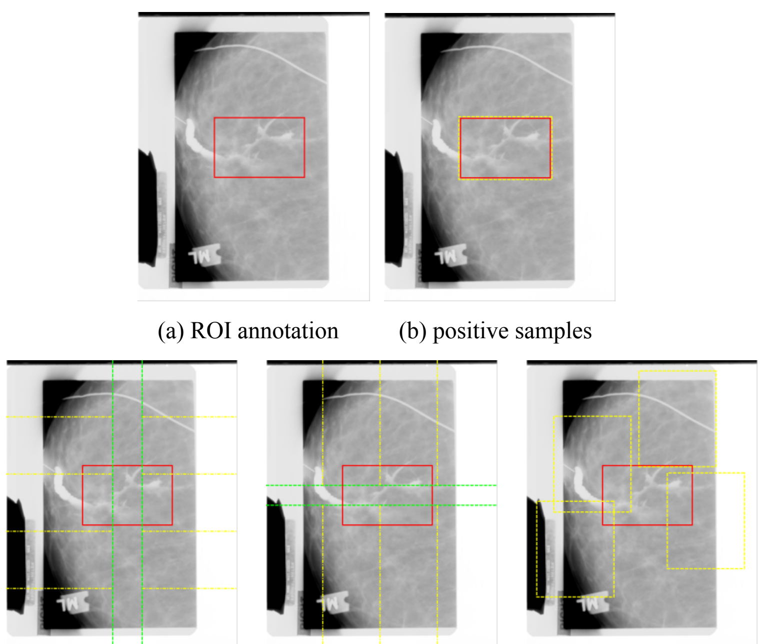
The ROI always contains vessels which have high contract against with the while image. The Haar features would extract such pattern to represent the ROI.

2.2 Generating training samples

One challenge in our task is the limited number of training samples. For this purpose, we boost both the positive and negative samples by generating more samples from the original expert annotations. Basically, additional positive samples are created by sampling uniformly around the annotated ROI; and negative samples are created by first decomposing an input image into several overlapped blocks and then randomly sampling from these blocks.

Since different expert could have different ROI annotations (as showed in Fig. 1), there is more flexible to provide ROI annotation for a require image. Given an annotated ROI with height h and width w, we also assume the region extended from original ROI as ground truth of ROI which with additional $\pm\Delta h$ in height and additional $\pm\Delta w$ in width. We could uniformly sample more regions from the extend area as positive training samples. The additional samples are with the same height and width as the original ROI. That means we'll have $\Delta h*\Delta h*\Delta w*\Delta w$ positive training instances included in the original ROI.

More negative training samples are also considered in our task. We first decompose the original image to four parts: up, bottom, left and right. The up and bottom parts only overlap the ROI with h/n_1 in height (see Fig .3(d), the region above the upper green line or below the lower green line). The left and right parts only overlap the ROI with w/n_2 in width. For each part, the part is divided into some patches with the same size. Negative training samples are extracted from the patches and the part itself. Meanwhile, we sample rectangles with certain size as the negative training samples in the whole original image. The overlapping parts of these rectangles and the ROI are less than a threshold both in width and height.



(a) ROI annotation (b) positive samples

(c) decomposition 1 (d) decomposition 2 (e) negative samples

Figure 3. Generating training samples from expert's annotation

In Fig.3 (a), the red rectangle is the ROI annotation by expert. Additional positive training examples with yellow (dash) color are showed in Fig.3 (b). We decompose the original image to up and bottom parts by green (dash) lines in Fig.3 (c). Green lines in Fig.3 (d) divide the original image into left and right parts. The up, bottom, left and right parts are decomposed into grids by yellow (dash dot) lines showed in both Fig.3 (c) and (d). Negative training examples are randomly selected from these parts and grids. We also extract negative training samples with certain size from the original image, which are showed in Fig.3 (e).

2.3 Merging and Filtering Detections

In the detection stage, by applying the trained boost classifier in (2) to an input image, usually many candidates will be detected. On the one hand, this is due to flexibility of positions of an ROI (e.g. Fig. 1), which means that many tightly overlapping regions can be almost equally good for serving as ROIs. On the other hand, due to the complex clutter structures in mammographic images, false positives are often detected.

We handle overlapping detected candidates through merging. The first case of overlapping is that there are some of candidate rectangles are slightly overlapped. Roughly speaking, let R={r_1, r_2,…, r_m} be a set of m tightly overlapping rectangles, the merging process generate a new rectangle s by taking the union of overlapping rectangles,

$$s = \cup_{i=1}^{m} r_i \tag{3}$$

Suppose that we have two rectangles r_1 and r_2 which represent the detection candidates. The heights and widths of the two rectangles are h_1, $w_1$and h_2, w_2, respectively. The criterion whether these two candidates should be merged into one is

$$r_2.x \leq r_1.x + r_1.width * m \quad r_2.x \leq r_1.x - r_1.width * m$$

$$r_2.y \leq r_1.y + r_1.width * m \quad r_2.y \leq r_1.y - r_1.width * m$$

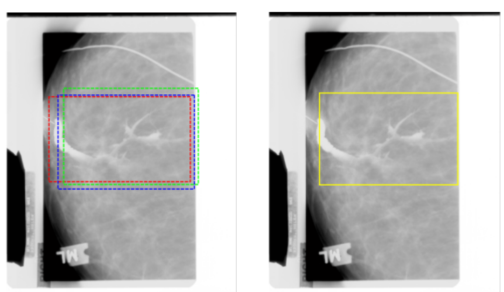
$$r_2.widht \leq r_1.width * (1+m) \quad r_1.widht \leq r_2.width * (1+m)$$

where, m is a parameter to control the merging process. If these two detection rectangles satisfy this merging criterion, then these two candidates will become to one detection rectangle. The average bounding box of these merged rectangles will be the final detection result.

After merging, there are usually a few detected candidates left. We can further filter out false detections by analyzing their “confidences”. Note that (2) gives confidence for each detection, we therefore need to “merge” confidences of overlapping detection as well. Denote $c(r)$ as the confidence of a detected rectangle, we investigate three schemes,

$$c_{sum}(s) = \sum_{i=1}^{m} c(r_i) \qquad c_{mean}(s) = \frac{\sum_{i=1}^{m} c(r_i)}{m} \qquad c_{max}(s) = max_{i=1}^{m} c(r_i)$$

to extend the confidence to the merged detection result. In other words, if several candidates are merged into one detection result, we use these different confidence score to evaluate the final detection results.



(a) before merging (b) after merging

Figure 4 Merging detection candidates

Fig. 4 (a) showed the detection candidates with different colors (dash). The merging detection result with yellow color is displayed in Fig.4 (b). The final detection result is the average bounding box of the detection rectangles in Fig. 4 (a). Meanwhile, the different confidence scores are calculated in the merging process.

Because there may be several detection results for each image, top ranks of candidates are recognized as detection result. In the experiment, several ranks are tested to filter false detections according to these different types of confidence scores.

3. EXPERIMENTS

3.1 Experiment Setup

We have used a dataset containing 28 galactographic images from 12 patients with expert annotated ROIs. The images were selected in an IRB approved study of existing anonymized galactographic images at the University of Pennsylvania. The mean age of the patients in that study was 44 years (range 27-64). The analyzed set of 28 galactograms included 12 images with no radiological findings, 2 images from malignant cases, 1 image with a benign finding; radiological findings were not available for 4 images. Experienced medical physicists from the University of Pennsylvania performed manual segmentation of ducts from clinical galactograms. The smallest rectangular ROIs covering all the manually segmented ducts were used for validation of the proposed automated ROI detection method."Some example images are shown in Fig. 1 and 4. We use four-fold cross validation for evaluating the proposed method. The data is randomly divided into four groups, each containing seven images. Then, for each fold, the seven images from one group are used for testing, and rest 21 images are used for training.

The method to extend training samples is utilized in the experiment. In order to get more positive training samples, we set $\Delta h = \Delta w = 3$. Therefore, for each training image, we could extract 81 positive images. We compose the image to four parts as discussed in Section 2.2. The up and bottom parts overlap with the ROI $h/3$ in height, which means that the parameter n_1=3. Parameter n_2 is set to 3 in the experiment and the left and right parts are overlapping with ROI $w/3$ in width. 6 images are randomly extracted as negative training samples from each part. By considering the size of the original image, we divide the up and bottom to 4 grids. Left and right parts of the original image are decomposed into 5 equalized grids. Each grid is recognized as a negative sample. For the reason that we want to cover all the possible regions as negative training samples, we also randomly choose negative samples with certain size from the whole original image. There are four different sizes are chosen in the experiment. We note that the original image size is *w_original* and *h_original* in width and height. The width for the certain size of the randomly region is *w_certain* and the height is *h_certain*. In the experiment, we set *w_ certain = w_original /3, w_ certain = w_original /2, h_ certain = h_original/3 and h_certain = h_original/2*. For each combination of *w_ certain* and *h_ certain,* we randomly extract 30 negative training samples with the certain size from the original image. The threshold in the randomly extracting approach is that the negative sample selected overlaps with ROI less than 1/2 of ROI. Finally we would sample 166 negative samples from one annotated image.

We have 81 positive and 166 negative training samples for only one original image with ROI annotation. In each cross validation, we have 81*21= 1701 positive samples and 166*21 = 3486 negative samples.

3.2 Evaluation:

After generating all the training samples, a classifier or detector could be obtained by AdaBoost. In the detection process, the parameter m of merging criterion is 0.2. For evaluation, the average detection rate over 4 different folds is used. In each fold, we have 21 images for training and 7 images for testing. The detection rates among the top1 to top 3 according to detection score are reported and we also compare the performances of the three false positive elimination schemes. The evaluation criterion could be written as:

$$jaccard\ coefficient = \frac{\cap \{ROI_{detection}, ROI_{groungdtruth}\}}{\cup \{ROI_{detection},\ ROI_{groungdtruth}\}}$$

where, the $ROI_{detection}$ is defined as the union of the overlap regions among the top 3 ROI detection candidates . If the detection results overlap with ground truth more than a threshold (the threshold in the experiment is 30%), we assume that the detection could be accepted. Hit rate or detection rate will be collected as the final result.

4. RESULTS AND DISCUSSION

In the experiment, a cascade classifier with 15 stages is obtained by using training samples for each cross-validation. We try to compare the three different method described above. The ROI detection results are summarized in Table 1. These results show clearly the effectiveness of the proposed method.

Table 1. Detection rates calculated using different filtering strategies applied in the 4-fold cross-validation experiment using 28 clinical galactograms:

Filtering Strategy	Top 1	Top 2	Top 3
Sum	75.00%	89.29%	96.43%
Mean	57.14%	82.14%	96.43%
Max	67.85%	96.43%	96.43%

The detection rate of the top 1 detection result is not good as if we consider top 2 or top 3 candidates detection result. It is reasonable because the images in the dataset are from different patients. The vibration of the pattern of ROI belongs to different patients are large. The might be caused by different vessel width, different intensity contrast or different vessel distribution in the ROI.

If we consider other detection candidates according to the detection score of different strategies, the detection rates are much better than only top 1 detection result is used. It also gives doctors choices to select truly ROI by combination with top candidates detection result.

Several examples of detections are displayed in Fig. 5. In these examples, the top 1 detection result well hit the annotation of expert. If top 1 result doest hit the groundtruth, we can see that the union of overlaps calculated by the top 3 candidates covers the annotation well (Fig. 6).

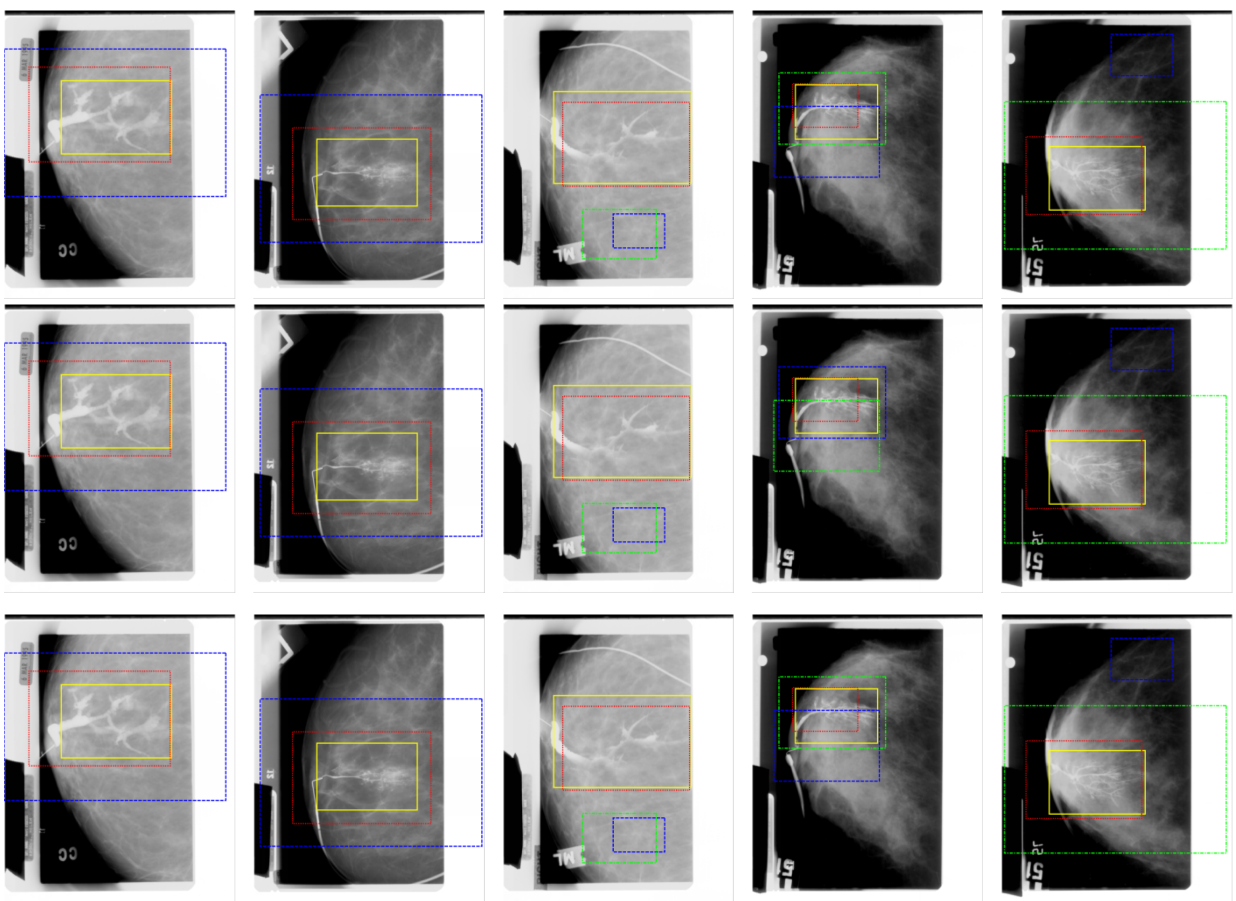


Figure 5. Example detections (top 1 result hits the groundtruth)

In Fig. 5, these rows are several detection results using sum, mean and max filtering method, respectively. The yellow (solid) rectangles are ROI annotation by expert. Top 1, 2 and 3 candidates are red (dot), blue (dash) and green (dash dot) rectangles. We can draw that the top first detection result is almost fit the expert annotation.

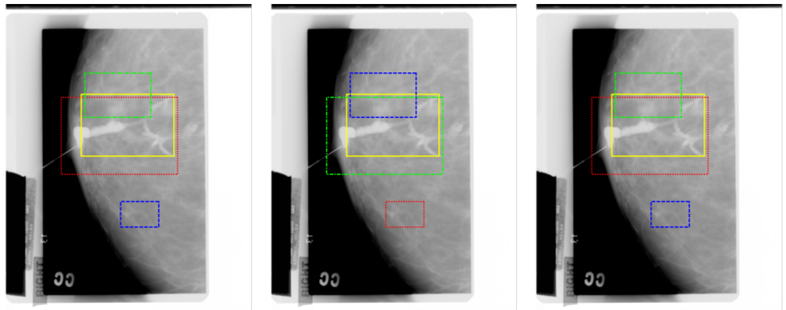


Figure 6. Example detections (union of top 3 candidates hit the groundtruth)
Merging methods from left to right: sum, mean and max

5. CONCLUSION

In this paper we propose using machine learning techniques to automatic detect ROIs from mammographic images. Using AdaBoost followed by domain adjusted post-processing such as false positive filtering, our approach achieved promising preliminary results. The study motivates us to further investigate the learning based solutions as well as using large datasets. We will also study related tasks such as vessel structure analysis and diagnosis.

6. ACKNOWLEDGEMENTS

This work was supported in part by NSF Research Grant IIS-0916624. The funding agency specifically disclaims responsibility for any analyses, interpretations and conclusion. We thank experts from University of Pennsylvania for providing data and the annotations.

REFERENCES

[1] Pereira, B., Mokbel, K., “Mammary ductoscopy: past, present, and future,” International Journal on Clinical Oncology, vol. 10, pp. 112-116 (2005).

[2] Megalooikonomou, V., Barnathan, M., Kontos, D., Bakic, P., Maidment, A.D.A., “A Representation and Classification Scheme for Tree-like Structures in Medical Images: Analyzing the Branching Pattern of Ductal Trees in X-ray Galactograms,” IEEE Transactions on Medical Imaging, Vol. 28, No. 4, pp. 487-493, (2009).

[3] H. Li, M.L. Giger, O.I. Olopade, A. Margolis, L. Lan, & M.R. Chinander., “Computerized Texture Analysis of Mammographic Parenchymal Patterns of Digitized Mammograms,” Academic Radiology, vol. 12, 863-873, (2005).

[4] Viola, P. and Jones, M., “Robust Real-Time Face Detection,” International Journal of Computer Vision, 57(2):137-154, (2004).

[5] Freund, Y. and Schapire, R. E., “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” J. Comput. Syst. Sci, 55(1): 119-139 (1997).

[6] Etta D. Pisano, Elodia B. Cole, Bradley M. Hemminger.e.al., “Image Processing Algorithm for Digital Mammography: A Pictorial Essay,” Radiographics, vol. 20, 1479-1491, (2000).

[7] Basim Alhadidi, Mohammad H. et al., “Mammography breast cancer image detection using image processing functions,” Information Technology Journal 6(2), 217-221, (2007).

[8] W. P. Kegelmeyer, Jr., J. M. Prineda, et al., "Computer-aided mammography screening for speculated lesions," Radiology, 191, 331-337, (1994).

[9] M. A. Kupinski and M. L. Giger., "Investigation of regularized neural networks for the computerized detection of mass lesions in digital mammograms," International Conference Engineering Medicine Biology, 1336-1339, (1997).

[10] H. Ling, M. Barnathan, V. Megalooikonomou, P. Bakic, A.Maidment., "Probabilistic branching node detection using local hybrid features," Proceedings of the 6th IEEE International Symposium on Biomedical Imaging (ISBI), (2009).

[11] T. Nuzhnaya, M. Barnathan, H. Ling, V. Megalooikonomou, P. Bakic, A.Maidment., "Probabilistic branching node detection using AdaBoost and hybrid local features," Proceedings of the 7th IEEE International Symposium on Biomedical Imaging (ISBI), (2010).

[12] G. Cardenosa, C. Doudna, G.W. Eklund, "Ductography of the breast: technique and findings," Am. J. Roentgenol., May 1994, Vol. 162, pp. 1081 - 1087.

[13] Bakic, P.R., Albert, M., Brzakovic, D., and Maidment, A.D.A.: "Mammogram Synthesis Using a 3D Simulation. III. Modeling and Evaluation of the Breast Ductal Network," Medical Physics 30: 1914-1925, 2003.

[14] M. Barnathan, J. Zhang, D. Kontos, P. Bakic, A. Maidment, V. Megalooikonomou: "Analyzing Tree-Like Structures in Biomedical Images Based on Texture and Branching: An Application to Breast Imaging", p. 25, Proc. IWDM 2008, Tucson, AZ, LNCS 5116, Springer-Verlag, Berlin, Germany.