MASS DETECTION IN DIGITAL MAMMOGRAMS BASED ON SELECTIVE SEARCH AND BAG OF WORDS

Abdullah Thabit, Khrystyna Faryna, Valeriia Abramova, Roa'a Khaled

University of Cassino and Southern Lazio, Cassino, Italy

ABSTRACT

Among the variety of abnormalities present in mammograms, it is crucial to detect masses and micro-calcifications. Mass detection is a challenging task due to diversity of mass sizes, shapes and poor contrast with surrounding tissue. In this work, we present a fully automatic mass detection method based on Selective Search algorithm. In the preprocessing stage Contrast Limited Adaptive Histogram Equalization (CLAHE) and Multi-scale Morphological Sifting (MMS) are utilized. Afterwards, the image patches with mass candidates are generated by Selective Search. The feature extraction is performed on original image patches as well as sifted. The classification of candidates into mass and non mass is performed through bootstrap aggregation with a bagging classifier. A FROC curve of the system was plotted, with the decision boundary threshold varying from 0.5 to 1. At a threshold of 0.5, it can attain sensitivity up to 80% , in the cost of false positive rate of about 10 per image.

Index Terms— Mass detection, selective search

1. INTRODUCTION

Breast cancer is one of the most common types of cancer in the world. Statistics says that in 2019, 30% of newly diagnosed cancers in women will be breast cancers [1]. At the moment, there is no method to prevent the disease but early detection can increase the chances for successful treatment [2]. This has motivated the development of many screening tools such as X-ray mammography, where it is considered to be the gold standard for screening and diagnosis of breast abnormalities.The most common abnormalities are micro-calcifications and masses, where masses are thought to be more challenging to be detected because of their large variation in size and shape, and also due to their poor contrast with surrounding tissues.

Over the past years, different methods have been proposed for mass detection in mammography. Some of them are based on region growing such as proposed in [3, 4]. The method is based on combining neighboring pixels by means of fulfilling a particular similarity criterion starting from some seed points. Some other methods are contour-based which aim to detect mass boundaries [5, 6]. Both of these approaches primarily segment mass candidates then classify them into mass objects and non-mass objects.

A different approach for object recognition that is quite popular in computer vision is to exhaustively search for objects candidates through a scanning window sliding over the image. This approach would first classify the candidates into their classes then segment their contours. Selective Search, a method proposed by Uijling et al. [7] follows a similar path but takes advantage of segmentation information to limit the search for proposals. It starts with an oversegmentation of an image, then it utilizes a hierarchical bottom-up grouping to gather regions until the whole image becomes a single region. The main goal of this method is not the perfect segmentation, but generating as many region proposals so that it is more likely that some of them have high overlap with true objects. To capture all possible objects, different combinations of several grouping criteria can be used such as colour, texture, size and shape [7]. Finally, all the object hypotheses from different variations of grouping algorithm are combined and ordered in a way that locations which are more likely to be an object come first.

Up to our knowledge, Selective Search (SS) has been only used in computer vision applications with colored images, but not with medical images. The goal of our work is to study the feasibility of using SS for mass detection in mammograms by proposing a system that is based on SS and Bag of Visual Words (BOVW). Fig.1 shows the pipeline of our system, which is further detailed in the following subsections.

2. MATERIALS AND METHODS

The system was trained and evaluated on the INBreast dataset [2]. This dataset is freely accessible. It contains 410 images with 115 mass cases in 107 images, of which 90 cases are from women with both breasts (4 images per case) and 25 cases are from mastectomy patients (2 images per case). Several types of lesions (masses, calcifications, asymmetries, and distortions) are included. The dataset also comes with the groundtruth annotations of the masses.

The implementation of the method was done in Python, taking advantages of external libraries, such as OpenCV, scikit-learn and Mahotas.

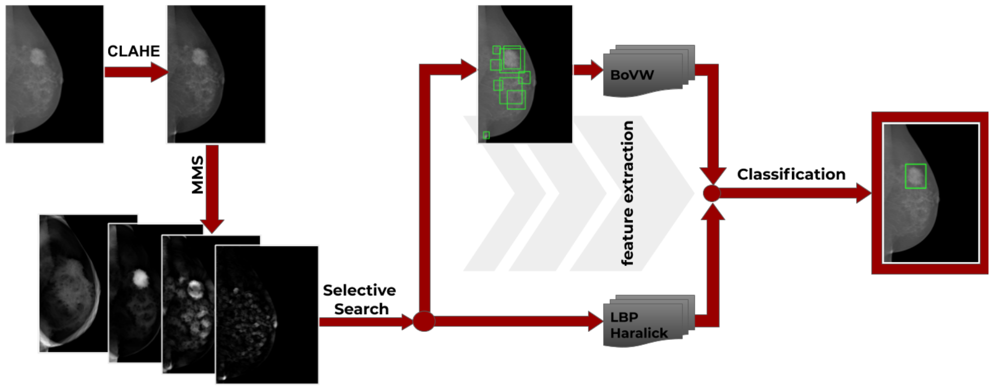

Fig.1 Pipeline of the proposed solution

2.1. Preprocessing

Mammograms usually suffer from low contrast, therefore a good preprocessing technique is crucial for mass detection. In this work, different preprocessing methods have been tested, including Contrast Limited Adaptive Histogram Equalization (CLAHE), which has been reported to work well with Mammograms [8], Mean-shift filtering, Rotational morphology [9] and Multiscale Morphological Sifting (MMS) [10].

For our mass detection system, MMS has been eventually adapted due to its superior performance. It is based on using sets of two linear structuring elements in different orientations. We iteratively apply a pair of structuring elements with the same orientation to an input image. As an output, we get an image, which is generated by normalizing the combination of all the result images. We consider that the objects we are looking for are composed with the same straight ”lines” in different sizes and directions as the structuring elements we use. Also, we want to catch masses of different sizes, therefore we apply the algorithm in multiple scales.

2.2. Candidates extraction

To extract the mass candidates, Selective Search as suggested in [7] was applied on the 4 morphological sifted scales of each image in the dataset. For the first graph segmentation of the algorithm, K was set to the default value (k=150), then using a bottom up approach, similar regions were grouped together based on multiple similarity measures including shape, size and texture. This result in many proposals (bounding boxes) of objects that have the potential to be masses.

Selective search is known to return proposals in the order of hundreds to thousands [7]. Therefore, limiting conditions were applied to eliminate the least probable proposals, leaving only the ones that are more likely to be masses. For instance, we excluded the regions near the breast boundary with a lot of background area; also, the size of bounding box was limited to the sizes of structuring elements at each scale.

Furthermore, the number of retrieved proposals for each object (mass) in the image depends on how likely for it to be an object. In other words, masses that have good contrast with the background have multiple bounding boxes in the retrieved proposals. These multiple proposals are kept to help fighting the class imbalance and were further eliminated at a later step.

2.3. Feature extraction

Since selective search is faster than other exhaustive search methods, with a much lower number of returned proposals, it can take the advantage of a more powerful and computationally demanding descriptor such as Bag of Visual Words (BOVW) [7]. Therefore, a BOVW model was constructed following the implementation in [11] as it gave more flexibility than the implementation in Opencv. The codebook of the model was set to have 100 clusters, and featuers were extracted from the contrast-enhanced images using a modified version of Harris detector GFTT [12] for key points detection and RootSIFT [13] for describing the key points neighborhood.

For robustness, other texture features were extracted from the multi scale sifted images, including 26 rotation invariant Local Binary Pattern (LBP) features [14] and 13 Haralick texture features from the Gray Level Occurance Matrix (GLCM) [15], including energy, entropy, contrast and homogenity.

2.4. Classification

Bootstrap aggregation with a bagging classifier was used to classify proposals into mass and non-mass objects. Bagging classifiers are known to perform best with algorithms that have high variance such as decision trees [16]. In this work, 200 decision trees have been used each of which uses 30% bootstrap sample of the training data, with a maximum depth of 3 for each tree. To account for the class imbalance, classes were weighted inversely proportional to how frequently they appear in the data. This way mistakes of classifying mass objects (minority class) are penalized more.

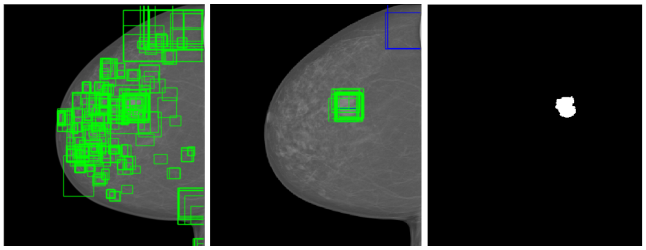

Fig.2 Candidate extraction, classification and post processing

Hard-negative mining was then applied to help reducing the false positives. It is based on retraining the classifier on hard-negative samples (negative samples that were classified as masses). These samples were ordered according to their prediction probability of being a mass, where the highest probable ones at the top, then taking only the top 25% of these samples to retrain the system.

Finally, a post processing step was then needed to eliminate the multiple detection of masses. First of all, from original image we extract patches, bounding boxes of which were labeled as masses. We threshold each of them with Otsu, select the largest connected component from a binary image, and sum the binary masks from each of the patches in a grayscale image. The final image is then thresholded again to create a binary mask.

A sample result of candidate extraction, classification and post processing is shown in Fig. 2.

3. RESULTS AND DISCUSSION

For system evaluation, Free-response Receiver Operating Characteristic (FROC) curve was calculated, where the FROC curve is defined as the plot of True Positive Rate (TPR) on the y-axis versus the rate of False Positives Per Image (FPPI) on the x-axis.

Fig. 3 shows our systems FROC curve for 30 points with the decision boundary threshold varying from 0.5 to 1. The figure shows that the system did not perform very well. at a threshold of 0.5, it can only attain sensitivity up to 80% but also in the cost of high false positive rate of about 10 per image.

Different attempts were tried to reduce the number of false positives including hard-negative mining, which helped in reducing it from 30 to 10 FPPI, however, more reduction of false positives always led to a decrease in sensitivity, which cannot be tolerated for a mass detection system. Other attempts included extracting features other than the originally adapted BOVW, where Local binary patterns and Haralick texture features increased the systems ability in discriminating between mass and non-mass objects but not very much.

One of the disadvantages of our mass detection system is that it is not based on segmentation, therefore access to features such as area, convexity, eccentricity and other shape features was not possible, where these features have been proven to work well for mass detection [2]. Another major limitation of using Selective search for mass detection is that it was developed for colored images with clear visual distinction between objects, therefore its performance with low contrast images did not work out well.

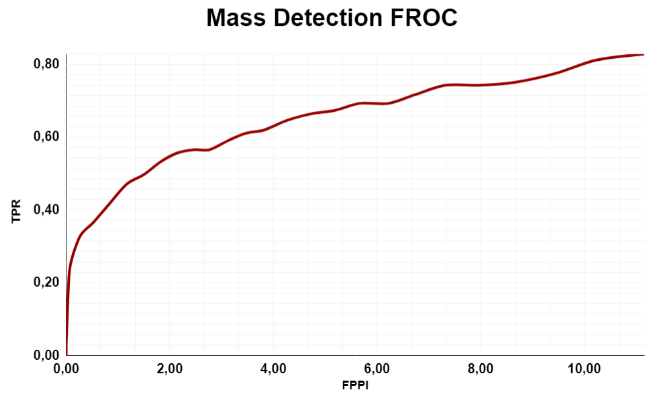

Fig.3 FROC curve of the proposed system

Table 1. Number of region proposals per method

Method	Masses detected (out of 115)	Total number of region proposals
CLAHE	88	6437
CLAHE+MeanShift	95	8136
Rotational morphology	101	40558
Morphological Sifting	115	74559
Morphological Sifting with extra conditions	115	29402

The result of applying selective search on the aforementioned preprocessing techniques is outlined in Table 1. It presents the number of detected masses along with the total number of retrieved proposals, where a mass was considered to be detected if its bounding box achieves a dice coefficient of 0.5 or higher with the bounding box of the ground truth.

As shown in Table 1, selective search was able to detect all masses only with MMS as a preprocessing step, which came at a cost of many retrieved proposals that required a lot of computational power. However, the major limitation of applying selective search on MMS is that the quality of the bounding boxes were affected. It resulted in some bounding boxes that were much bigger that the actual mass, while some others were too small to contain the whole mass region (Fig. 4). These bounding boxes led to a confusion for the classifier as the features coming from them were similar to those containing no masses.

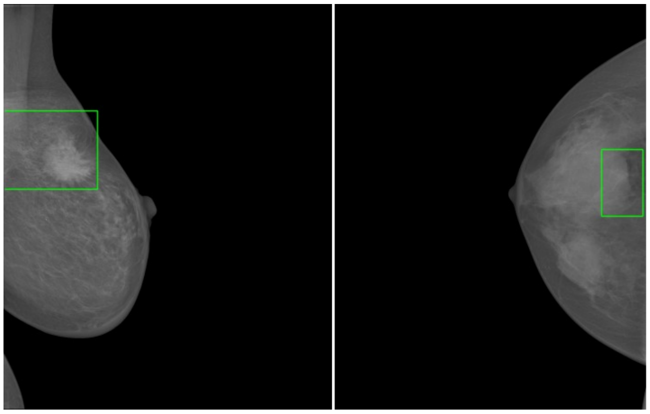

Fig.4 Bounding boxes bigger/smaller than actual mass

4. CONCLUSION

Despite the good performance of Selective Search for object detection in computer vision, It appeared to be not suitable for medical imaging applications, such as in mammography, where objects are not easily distinguishable from their background. Even with computationally heavy preprocessing techniques such as Multiscale Morpholgical Sifting, the number of retrieved bounding boxes and their quality have overcomplicated the problem without producing any substantial gain in performance. Further improvement of the proposed system could take place by investigating other preprocessing techniques and features. However, the extent for improvement of this algorithm appeared to be limited.

5. REFERENCES

[1] Siegel R.L., Miller K.D., and Jemal A, "Cancer statistics, 2019," *CA Cancer J Clin*, vol. 69, pp. 7–34, 2019.

[2] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: Towards a full field digital mammographic database," *Academic Radiology*, vol. 19, pp. 236–248, 2012.

[3] S. Punitha, A. Amuthan, and K. Suresh Joseph, "Benign and malignant breast cancer segmentation using optimized region growing technique," *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 348 – 358, 2018.

[4] Nevine H Eltonsy, Georgia Tourassi, and Adel Elmaghraby, "A concentric morphology model for the detection of masses in mammography," *IEEE transactions on medical imaging*, vol. 26, pp. 880–9, 07 2007.

[5] Shi Jiazheng, Sahiner Berkman, Chan Heang-Ping, Ge Jun, Hadjiiski Lubomir, A Helvie Mark, Nees Alexis, Yi-Ta Wu, Jun Wei, Chuan Zhou, Yiheng Zhang, and Jing Cui, "Characterization of mammographic masses based on level set segmentation with new image features and patient information," *Medical physics*, vol. 35, pp. 280–90, 02 2008.

[6] Donato Cascio, F Fauci, R Magro, G Raso, Roberto Bellotti, F De Carlo, Sabina Tangaro, Giorgio De Nunzio, Maurizio Quarta, Giorgio Forni, Adele Lauria, Maria Fantacci, Alessandra Retico, Giovanni Masala, P Oliva, Stefano Bagnasco, Sorin Cheran, and E Lopez Torres, "Mammogram segmentation by contour searching and mass lesions classification with neural network," *Nuclear Science, IEEE Transactions on*, vol. 53, pp. 2827 – 2833, 11 2006.

[7] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, and A.W.M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, 2013.

[8] D Narain Ponraj, M Evangelin Jenifer, P Poongodi, and Manoharan Samuel, "A survey on the preprocessing techniques of mammogram for the detection of breast cancer," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, 01 2011.

[9] Yoshitaka Kimori, "Mathematical morphology-based approach to the enhancement of morphological features in medical images," *Journal of clinical bioinformatics*, vol. 1:33, 2011.

[10] Hang Min, Shekhar S. Chandra, Stuart Crozier, and Andrew P. Bradley, *Breast Mass Segmentation based on Multi-scale Morphological Sifting and Self-grown Cascaded Random Forests*, 2012.

[11] Adrian Rosebrock, "The bag of (visual) words model," https://www.pyimagesearch.com/pyimagesearch-gurus/.

[12] Jianbo Shi and Carlo Tomasi, "Good features to track," *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 600, 03 2000.

[13] Relja Arandjelovic and Andrew Zisserman, "Three things everyone should know to improve object retrieval," 01 2012.

[14] Timo Ojala, Matti Pietikinen, and David Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, pp. 51–59, 01 1996.

[15] S. Ohmshankar and C. K. Charlie Paul, "Haralick fetaures based mammogram classification system," in *Second International Conference on Current Trends In Engineering and Technology - ICCTET 2014*, July 2014, pp. 409–413.

[16] Leo Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123–140, 08 1996.