Computers in Biology and Medicine 110 (2019) 244–253

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Mammogram segmentation using multi-atlas deformable registration

Manish Kumar Sharma[a], Mainak Jas[b,c], Vikrant Karale[a], Anup Sadhu[d,e], Sudipta Mukhopadhyay[a,*]

[a] *Department of Electronics and Electrical Communication Engineering, IIT Kharagpur, India*
[b] *Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, USA*
[c] *Department of Electrical Engineering, IIT Kharagpur, India*
[d] *MRI Scan Center, Medical College, Kolkata, India*
[e] *EKO X-Ray and Imaging Institute, Kolkata, India*

ARTICLE INFO

Keywords:
Mammograms
Breast region segmentation
Clustering
Atlas based image registration
Atlas selection

ABSTRACT

Accurate breast region segmentation is an important step in various automated algorithms involving detection of lesions like masses and microcalcifications, and efficient telemammography. While traditional segmentation algorithms underperform due to variations in image quality and shape of the breast region, newer methods from machine learning cannot be readily applied as they need a large training dataset with segmented images. In this paper, we propose to overcome these limitations by combining clustering with deformable image registration. Using clustering, we first identify a set of atlas images that best capture the variation in mammograms. This is done using a clustering algorithm where the number of clusters is determined using model selection on a low-dimensional projection of the images. Then, we use these atlas images to transfer the segmentation to similar images using deformable image registration algorithm. Our technique also overcomes the limitation of very few landmarks for registration in breast images. We evaluated our method on the mini-MIAS and DDSM datasets against three existing state-of-the-art algorithms using two performance metrics, Jaccard Index and Hausdorff Distance. We demonstrate that the proposed approach is indeed capable of identifying different types of mammograms in the dataset and segmenting them accurately.

1. Introduction

Automatic segmentation of Breast Region is a critical step in Computer Aided Diagnostic (CAD) tools for mammograms – with applications including detection of lesions, classification of lesions and diagnosis of breast cancer. It enables algorithms downstream in the CAD pipeline to be more efficient and accurate by constraining the analysis to be confined to a region of interest relevant to the algorithm.

Multiple attempts at breast region segmentation have been made using thresholding [1] based techniques [2–9]. These methods estimate the intensity level which best separates the background from the foreground. Although these algorithms are computationally efficient, medical images are rarely strictly bimodal thus violating the fundamental assumption of these algorithms. Furthermore, thresholding algorithms do not take into account the relationship amongst pixels. Localized processing [4,8], region growing [5], fast-marching algorithm [10], connected component labeling [7], filter based pre-processing [6,9] are some of the methods with which the researchers have tried to overcome the limitations of thresholding. Pattern analysis and clustering based techniques [11–13] have attempted to automatically classify pixels as either belonging to the foreground or the background. More recently, deep learning based techniques [14–17] have also been applied to segment breast region. These methods require a large number of segmented training images which are not always available for mammograms. Deformable methods comprise another category of segmentation algorithms. These offer finer accuracy by taking into account the shape of the segmented region by deforming contours or objects while minimizing an energy functional. Atlas-based segmentation [18], active contours [19–21], snakes [22] and level sets [23] are examples of this approach. Apart from these broad categories, several other attempts have also been made to identify the breast region boundary in mammograms. Karnan et al. [24] used a genetic algorithm based approach to identify the exact orientation of images and then used bilateral subtraction of left and right images. The watershed algorithm [25] is another approach where the image is treated as a 3D topography and the watershed lines separating neighboring basins are

---

[*] Corresponding author.

E-mail addresses: manishks344@gmail.com (M.K. Sharma), mainakjas@gmail.com (M. Jas), vicky753@gmail.com (V. Karale), sadhujee1@gmail.com (A. Sadhu), smukho@ece.iitkgp.ac.in (S. Mukhopadhyay).

https://doi.org/10.1016/j.compbiomed.2019.06.001
Received 11 February 2019; Received in revised form 17 May 2019; Accepted 3 June 2019

used to segment the image. Shi et al. [26] segmented breast region using sharp changes in gradient map of mammograms. Saidin et al. [27] proposed a graph-cut technique where the user needs to select a few pixels from the foreground and background each. A brief overview and comparison of the segmentation methods is provided by Mustra and Grgic in Ref. [28].

Despite the fact that image segmentation is a well-studied problem for several years, breast region segmentation still poses a challenge due to technical and physiological reasons. In this article, we will consider atlas-based image segmentation – a technique successfully applied in several imaging domains and modalities – neuroimaging [29], cardiac images [30], pulmonary imaging [31] *etc*. However, this has not yet been widely explored for breast region segmentation in mammograms. Jas et al. [18], in a previous conference paper, proposed an atlas-based breast region segmentation where landmark-based rigid registration followed by deformable image registration was applied. However, such a method crucially depends on the accuracy of the landmarks detected and the initial distance between the two images used for registration. In the case of mammography, automated landmark detection algorithms can detect at best a few points: the nipple [32], pectoral muscle edge [33], etc. Thus, a fully automated method is not feasible. Particularly because the breast tissue type varies considerably from individual to individual. The mini-MIAS [34] database, for instance, classifies the images into dense, glandular or fatty images. Thus, a strategy based on registration with a single atlas will most likely fail in this case. The tissue type (dense, glandular, or fatty) which predominately constitutes the breast region may affect the contrast of the image. The shape and size of the breast region also vary from individual to individual. A strategy based on multiple atlases instead of a single atlas will be more successful in such a scenario. Therefore, in this paper a multi-atlas based approach has been proposed for breast region segmentation.

In a multi-atlas strategy, one could either fuse segmentations obtained with different atlas images or generate one segmentation corresponding to the best atlas. Indeed, the selection of an optimal atlas has been demonstrated to affect the final segmentation accuracy by Wu et al. [35]. In this paper, we focus on the latter approach. We demonstrate that generating a database of atlases in a multi-atlas framework helps generate better segmentation masks. We formulate an automated breast region segmentation framework which uses two steps: first, selection of atlas and second, registration process and segmentation. We propose to create a data-driven library of atlas images. This is done by first identifying the types of images in the database and then selecting an optimal number of atlas images based on this information. Given a new image, its nearest matching atlas image is found from the set of candidate atlases and used for registration. The computed segmentation results are evaluated against ground-truth segmentation masks verified by an expert radiologist. We compare this approach against three competing methods on standard evaluation metrics and find that the proposed approach is able to outperform these by a large margin.

2. Methods

The proposed method involves two steps: first, selection of atlas and second, registration process and segmentation.

2.1. Atlas selection

The atlas selection procedure is data-driven. The efficiency and accuracy of the registration process depends on the atlas image used. Hence, instead of selecting atlas images randomly, we use clustering to determine which images will be the most representative ones. To efficiently compute the clusters as well as for visualization, we start by projecting the high-dimensional images to a 2D plane using t-distributed Stochastic Neighbor Embedding (t-SNE) [36]. The t-SNE method claims to preserve the inter-class separability while projecting to lower dimensions.

We fitted Gaussian Mixture Models (GMM) with k Gaussian distributions, $k \in [2..9]$, to this 2D projected data. The range of k can be configured depending upon the size and nature of the dataset. However, we prefer to select a small number of atlas images for segmenting, thus k should not be too large. The optimal value of k will give the optimal number of clusters k^*. Bayesian Information Criteria (BIC) [37] is a popularly used technique in model selection. Given a set of observations with sample size n, the number of images in the dataset, each GMM has a maximized value of likelihood function, $\hat{L}_k$. The BIC criteria in Equation (1) uses this value and the dimensionality of the GMM (the number of Gaussians), k, to find the most optimal model. The dimensionality, k, is used as a penalty term so that the model with highest dimensions is not always preferred. The model with minimum value of BIC gives the optimal number of clusters, k^*, Equation (2).

$$BIC_k = -2 \times \ln \hat{L}_k + k \times \ln n \tag{1}$$

$$k^* = \underset{k}{\operatorname{argmin}}\, BIC_k \tag{2}$$

We used the value of k^* obtained above to perform a k-means clustering on the actual images. The next step is to identify an atlas image from each cluster. Since the central cluster image is the average of all images belonging to that cluster, it is blurred and thus not suitable for registration. We therefore picked the image closest to the cluster center as the best representative atlas image for that cluster. The image closest to the cluster center C_k is the one which has the lowest Euclidean distance from C_k. Given images $\{I_{k1}, I_{k2}...\}$ belonging to the kth cluster, each with dimensions $W \times H$, the corresponding candidate atlas, $I_{atlas,k}$, is selected as given by Equation (3).

$$I_{atlas,k} = \underset{I \in \{I_{k1}, I_{k2}...\}}{\operatorname{argmin}} \sqrt{\sum_{j=1}^{WH} (C_k^j - I^j)^2} \tag{3}$$

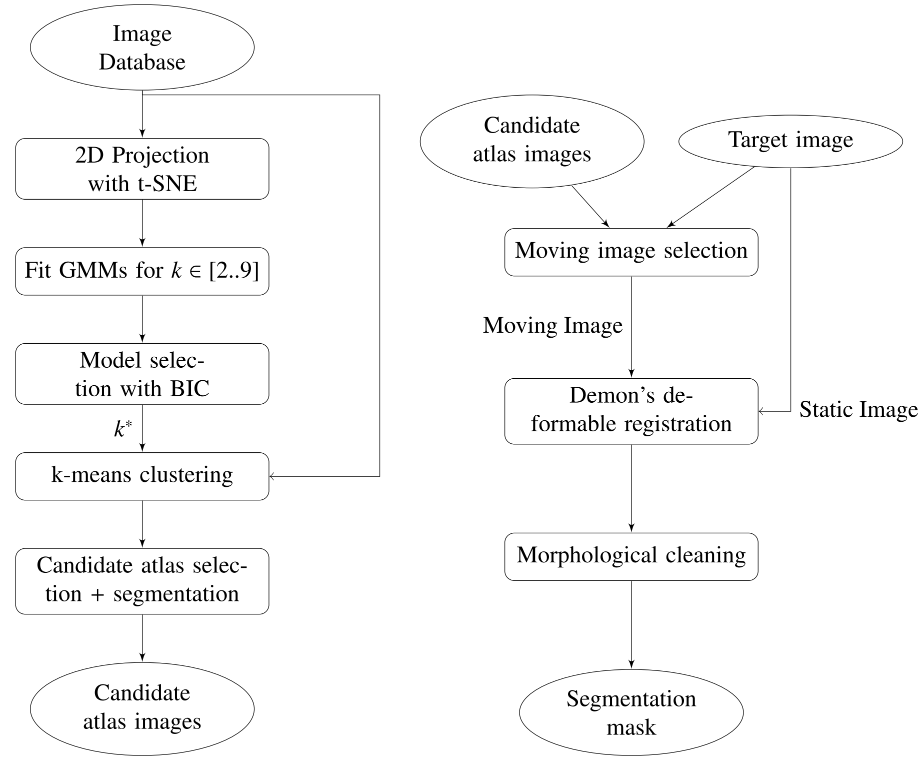
The candidate atlas images should be accompanied by ground-truth masks just as in any atlas-based segmentation procedure. This step is typically manual even though the segmentation procedure is entirely automatic. For the convenience of the reader, the atlas selection procedure is summarized in Fig. 1a.

2.2. Demon's registration

The candidate atlases are now used for the registration and segmentation. We used Demon's Deformable registration [38] for the registration process. The complete procedure for the registration is shown in Fig. 1b. Demon's registration requires two images: a static image, and a moving image. The moving image is transformed to match the static image during the registration procedure. As opposed to affine registration where the same transformation is applied on the entire image, in deformable registration, each pixel in the image can be transformed independently. Given a new image for segmentation, we consider it as the static image, S, and need to find the corresponding moving image, M. Since, we already have a set of candidate atlases from the previous step, we select the best matching atlas from this set as the moving image, M, for the registration procedure. In other words, we must decide which cluster $\hat{k}$ the image to be segmented belongs to, and assign the atlas from that cluster as the moving image. That is to say, the moving image $M = I_{atlas,\hat{k}}$ is selected using:

$$\hat{k} = \underset{k}{\operatorname{argmin}} \sqrt{\sum_{j=1}^{WH} (I_{atlas,k}^j - S^j)^2} \tag{4}$$

During the registration procedure, the pixels of the moving image are moved as per the accelerated Demon's equation given by Wang et al. [39] in Equation (5). For a moving image, M, and a static image, S, the additional transformation field, $\Delta\mathbf{U}^{(i)}$, for movement of intensity pixels in iteration i is calculated as



(a) Creating the set of candidate atlas images (b) Atlas based segmentation of new image

Fig. 1. Schematic flowchart for atlas-based segmentation with automatic atlas selection.

$$\Delta \mathbf{U}^{(i)} = (M - S) \times \left(\frac{\nabla S}{|\nabla S|^2 + \alpha^2 (S - M)^2} + \frac{\nabla M}{|\nabla M|^2 + \alpha^2 (S - M)^2} \right) \tag{5}$$

where α is the normalization factor that allows the transformation field strength to be adjusted adaptively in each iteration and ∇ computes the gradient for 2D images. Note that the transformation field $\mathbf{U}^{(i)}$ is a vector for each pixel with both an x and a y component.

The total transformation field after the ith iteration, $\mathbf{U}^{(i+1)}$, is the sum of previous transformation field and the additional transformation field calculated in iteration i scaled by a value of s_k, speed factor used to speed up the registration process. Equation (6) shows the transformation field update step.

$$\mathbf{U}^{(i+1)} = \mathbf{U}^{(i)} + s_k \Delta \mathbf{U}^{(i)} \tag{6}$$

Within each iteration, the total transformation field, $\mathbf{U}^{(i)}$, is regularized using Gaussian smoothing and the pixels of the moving image are translated with the smoothed transformation. The registration algorithm terminates when we meet the stopping criteria based on correlation of static and moving images given in Equation (7). Here, $\rho(S, M)^{(i)}$ is the 2D cross-correlation value of the static and moving images in iteration i. We measured the gain in cross-correlation value of static and moving images over ten iterations and stopped the registration process when this gain is smaller than a threshold τ.

$$\Delta \rho^{(i)} = \rho(S, M)^{(i)} - \rho(S, M)^{(i-10)} < \tau \tag{7}$$

After the convergence of registration process, the segmentation of the moving image is transferred to the static image.

2.3. Post-processing

Deformable registration is a pixel-level translation process. This often leads to the formation of holes in the mask. We use morphological cleaning operations, opening and hole-filling, using a disk-shaped structuring element to remove such holes.

3. Results

We experimented with two publicly available datasets for evaluation of our method — mini-MIAS and DDSM. We chose mini-MIAS dataset as our primary source and present both qualitative as well as quantitative results for the dataset. After verifying the efficiency of our algorithm on this dataset, we verified our approach using a subset of the DDSM dataset and present qualitative results over the images from DDSM dataset. The following subsections present results for the mini-MIAS dataset and the results on DDSM dataset are presented in Section 3.8.

3.1. Evaluation strategy

We chose mammographic images from the mini-MIAS dataset as our primary data source. The dataset consists of 322 images from 161 patients, with left and right mammogram pairs per patient. These mammograms were digitized at $50\mu m$ pixel edge and downsampled to $200\mu m$ pixel edge. We downsampled the original 1024×1024 images to 512×512 for ease of processing.

To standardize the images, we performed a series of preprocessing steps. First, we flipped the left breast mammograms so that all mammograms had the same orientation. Images in mini-MIAS dataset are padded on the left and right to make the images square. We removed this padding and translated the breast region so that it touches the right edge. These two preprocessing steps ensures that the images are in a common coordinate frame.

We wanted to evaluate how well our method generalizes when there are new images that are not observed during the atlas selection procedure. To do so, we used subject-wise 5-fold cross-validation. In one fold, four partitions of the dataset were used to identify the candidate set of atlas images and the fifth partition was used for segmentation. The subject-wise cross-validation strategy ensures that left and right mammogram images from the same subject were in the same fold. This way, there is no leakage of data between the training and the test set as it ensures that the same subject is *not* in both the training and test set. Indeed, this is to avoid inflated results as could be the case with a simple 5-fold cross-validation strategy [40].

For evaluating our results, we used ground truth segmentation masks prepared manually. The researcher MKS marked the boundaries manually on an enlarged image using an image editing software and then converted the boundary to mask. He overlaid the masks on actual images and got it verified by the radiologist AS. All ground truth masks were prepared at the resolution of original images, 1024 × 1024. For the benefit of the broader research community, we have made these masks publicly available[1].

All computations are performed using MATLAB R2014a running on a 64-bit Windows 8 machine with 4 GB RAM and *Intel (R) Core (TM) i5-2430 M CPU @ 2.*40 GHz Processor.

3.2. Atlas selection

We used t-SNE projection only to identify the number of clusters, k^* and not for segmentation. Hence, we used low-resolution images for this step to reduce unnecessary computational complexity. We downsized the 512 × 512 images to 32 × 32 resolution and then vectorized it to get a feature vector of length 1024. This feature vector was finally used for the t-SNE projection. The t-SNE algorithm [36] first reduces the dimensionality of feature vectors using Principal Component Analysis (PCA) and then computes the 2-D projections using these reduced dimensionality feature vectors. We have reduced the dimensionality of the feature vector to 25 × 1 after PCA. Here, we present the results from one fold of the cross-validation for illustrative purposes. However, in Section 3.7 when comparing against other algorithms, we aggregate the results across cross-validation folds.

Fig. 2a shows the t-SNE projection of the mammograms. The corresponding BIC of the GMMs with $k \in [2..9]$ is shown in Fig. 2b. The minima of the BIC curve gives the value of $k^* = 5$. The 2D projection of the mammograms in Fig. 2a shows that the mammograms do in fact belong to 5 different clusters. Fig. 3 shows images belonging to the five different clusters. Each row contains images from the same cluster and different rows belong to different clusters. It can be seen that while images belonging to the same cluster are similar in shape and intensity profile, images belonging to different clusters are notably different. It is quite easy to conclude from this figure that performing deformable image registration between images in the same cluster is far easier compared to images from different clusters.

We identified one candidate atlas from each cluster as the image closest to the central cluster image in terms of the Euclidean distance. Fig. 4 shows the selected atlas images for this cross-validation fold. The atlas images represent different types of breast images in the dataset. Hence, we argue that data-driven identification of breast shapes and tissue types play an important role in the segmentation process.

In order to segment the images in the dataset, we expect our atlas images to be already segmented as an input to the registration algorithm. Hence, we expect the user to provide the segmentation for this small subset of images identified as atlas images. This can be done using manually generated ground truth masks or some other algorithm which best suits the segmentation of individual images. In our case, since we already generated ground truth masks for verification of results, we used the same masks for segmenting the atlas images.

3.3. Mammogram segmentation

We performed the registration using 512 × 512 resolution images. To speed up the registration process, we chose the speedup factor of $s_k = 4$. The average time for the overall registration process was 70.06 s per image. The average number of iterations for convergence was 393 per image. The maximum number of allowed iterations was 800.

We used the difference in the 2D correlation coefficient of static and moving images as stopping criteria in Equation (7). We chose the threshold for stopping, $\tau = 0.001$. Normalization factor, α, allows force strength to be adjusted adaptively during iterations [41]. We chose it to be 1, as in Thirion et al. [38], to remove its impact and simplify the equation. The value of the standard deviation for Gaussian smoothing, σ, allows the contours to be curvy, a very small value may result into wiggly contours while a large value may cause the contours to be too smooth. We chose $\sigma = 30$ as suggested by Jas et al. in Ref. [18].

3.4. Post-processing

Since deformable registration is a pixel-level translation process, it leads to the formation of holes in the mask. We used a disk-shaped structuring element of radius 10 pixels for refining the obtained masks. Also, we observed that some holes were formed along the edges of the image. Recall that in the data preparation stage, we aligned the breast region to the right and top edges. Thus, to recover these regions, we performed a hole-filling morphological operation along the right and top edges. Fig. 5 shows a segmentation mask before and after the post-processing step. Finally, for visualization purposes, the segmentation mask was converted into a contour using a Prewitt edge detector and was overlaid on the actual image along with the ground truth.

3.5. Segmentation results for different tissue types

The mini-MIAS dataset classifies breast tissues into three categories: Dense Glandular, Fatty, and Fatty Glandular. Fig. 6 shows the segmentation results for different types of tissues present in the mini-MIAS dataset. Each tissue type presents a different shape and intensity profile. Indeed, from visual assessment, we can conclude that the segmentation performs reasonably well irrespective of the tissue type. In Fig. 7a and b (described in Section 3.7), we also show that our method outperforms other methods for different tissue types. This can be partly attributed to the fact that the atlases learned in our method represent the variation in the breast shapes across tissue types.

3.6. Evaluation metrics

For evaluating the results, we used Jaccard Index [42] and Hausdorff Distance [43] as the accuracy metrics. Jaccard Index measures the similarity between two sets, A and B, as the ratio of the number of elements in the set intersection over the number of elements in the set union given by Equation (8).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{8}$$

Here, $|\cdot|$ gives the number of elements in the set, also known as the cardinality of the set. In the current context, the set of foreground pixels in the computed segmentation masks and ground truth masks are A and B respectively. Finally, cardinality of the set is defined as the number of logical ones present in the mask. A higher value of Jaccard index indicates good segmentation while lower values indicate poor segmentation. A value of 1 would indicate perfect segmentation.

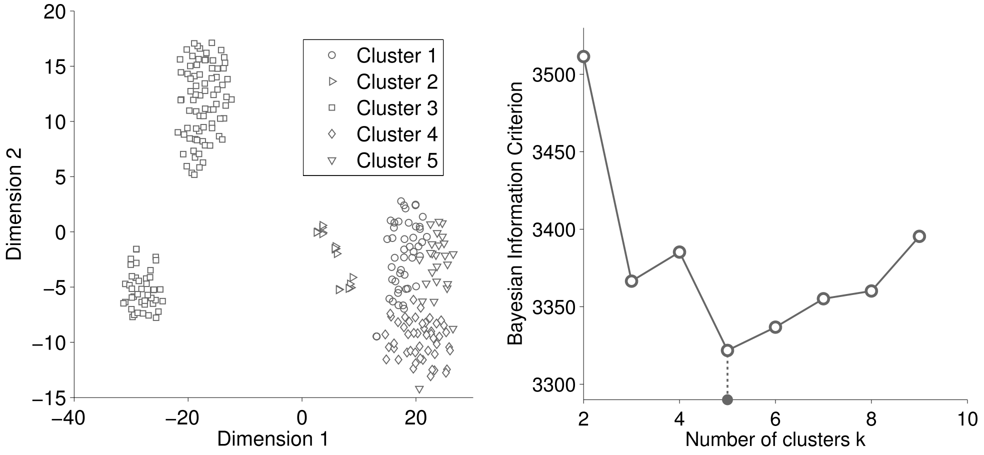
The second metric, Hausdorff distance, measures the maximum of the distances between points of one set to corresponding nearest point in another set. For $a \in A$ and $b \in B$, the Hausdorff distance, $h(A, B)$, is given by Equation (9), where $d(a, b)$ is the Euclidean distance between the points a and b.

$$h(A, B) = \max\left[\max_{a\in A}\min_{b\in B} d(a, b),\ \max_{b\in B}\min_{a\in A} d(a, b)\right] \tag{9}$$

As it measures distances rather than the number of elements, it is affected by the shape of the contours more than the area of the region. Thus, it is a metric complementary to the Jaccard Index. The lower the

---

[1] https://brsegmentation.github.io/



(a) t-SNE projection of mammogram images to 2 dimensions.

(b) Bayesian Information Criterion (BIC) of GMMs vs. k, the number of Gaussians

Fig. 2. 2-D projection of images from one fold of mini-MIAS dataset and corresponding BIC curve. The minima of the BIC curve suggests the presence of $k^* = 5$ clusters. The distribution of five clusters is shown in the projected 2-D space.

Hausdroff distance, the better it is. In our analysis, we used the set of coordinates of edge points belonging to the segmentation mask and ground truth mask as A and B, respectively. The edge points were detected using the Prewitt edge detection method.

3.7. Comparison against other algorithms

We benchmarked our algorithm against three state-of-the-art algorithms – Ferrari et al.'s Active Contour Model (ACM) [20], Arifin et al.'s Multi-level Thresholding (MTH) [2] and Abubaker et al.'s Row-by-Row Thresholding (RRT) [3]. Ferrari's method first approximates a breast boundary using a chain code for binarized images. This chain code is used as an input to an active contour model for identification of true breast boundary. Arifin's method of multi-level thresholding divides the histogram into different classes and binarizes the image at the interface of two clusters. Abubaker's row-by-row thresholding method identifies different thresholds for each row of the image. Each row is binarized using the threshold identified for that row to give the final segmentation. We generated segmentation masks using these methods.

Fig. 7a shows the average Jaccard Index and Fig. 7b shows the average Hausdorff Distance for different type of tissues present in mini-MIAS dataset. It can be seen from the figure that the performance of the proposed approach is consistently better than the competing algorithms. For each tissue type, the Jaccard Index for our method is higher compared to other methods, while the Hausdorff Distance is lower.

The quality of segmentation achieved by the proposed method and competing methods can also be visually compared in Fig. 8. We have marked the segmentation boundaries achieved by these methods and the segmentation boundary of the ground truth in the same image. The red curve represents the segmentation boundaries obtained by the algorithms while the blue curve represents the ground truth. For a fair assessment, Fig. 8 considers four different scenarios: 1. The proposed method outperforms competing methods, 2. All methods perform well, 3. Competing methods outperform the proposed method, and 4. None of the methods perform well. It is worth noting that though in some cases, the proposed method is not able to perform well, yet, the segmentation achieved is close to the actual ground truths.

We compared the performance of these methods quantitatively using the same ground truth masks that were used for visual assessment. For multi-level thresholding and row-by-row thresholding method, the outputs were in the form of masks. However, for adaptive contour model method, the output was the contour of the breast region boundary. Hence, we do not report the Jaccard Index for adaptive contour model. For computing Hausdorff distance, we used the contour for the ACM method and the edges of masks for MTH and RRT. Table 1 summarizes the evaluated results for different segmentation methods. We can confirm that the proposed method's performance is certainly better than that of the competing methods in most cases for both the evaluation metrics. Indeed, the average Jaccard Index for the proposed method is the highest while the average Housdorff distance is the lowest. In fact, the proposed method not only performs well on average but also wins against the competing method in a large fraction of images: 79.50% for Jacard Index, and 74.22% for Hausdorff distance.

3.8. Extension to DDSM dataset

We extended the testing of the proposed algorithm to the DDSM database. We chose 532 images from 133 patients, where 4 images were taken per patient corresponding to the four views, left CC view, right CC view, left MLO view and right MLO view. We reduced the image bit depth to 8 bits from the original 16 bits to ease the computational burden. To standardize the images, we flipped the left breast mammograms and resized all images to a fixed height of 512 pixels. Then, zero padding was added to the left of the images to make the final width of images 462 pixels. Due to the large number of images, we couldnt create masks for this dataset and hence, present only the qualitative results.

Fig. 9a shows the t-SNE projection of the images from the DDSM dataset and Fig. 9b shows the corresponding BIC of the GMMs. The minima of the BIC curve, in this case, gives $k^* = 6$ which is used for clustering as shown in the t-SNE projection. Fig. 10 shows the segmentation of an image from the DDSM dataset with different methods. It can be seen that the segmentation achieved with the proposed method is indeed fairly accurate as compared to other methods.

4. Discussion

We discussed in the Introduction that though there are several approaches to segment breast region in mammograms, due to high diversity in the shape and intensity profile of mammograms (Fig. 3), an accurate segmentation is hard to achieve. Although Wirth et al. [44] demonstrate the effect of databases on the algorithm performance, most existing algorithms [2–13,18–27] segmented images without considering the variation in image shapes across the database. We have shown in this paper that our data-driven approach outperforms three state-of-the-art methods. The reason for this is that our method takes into account the shape and profile of breast tissue in mammograms over the dataset.

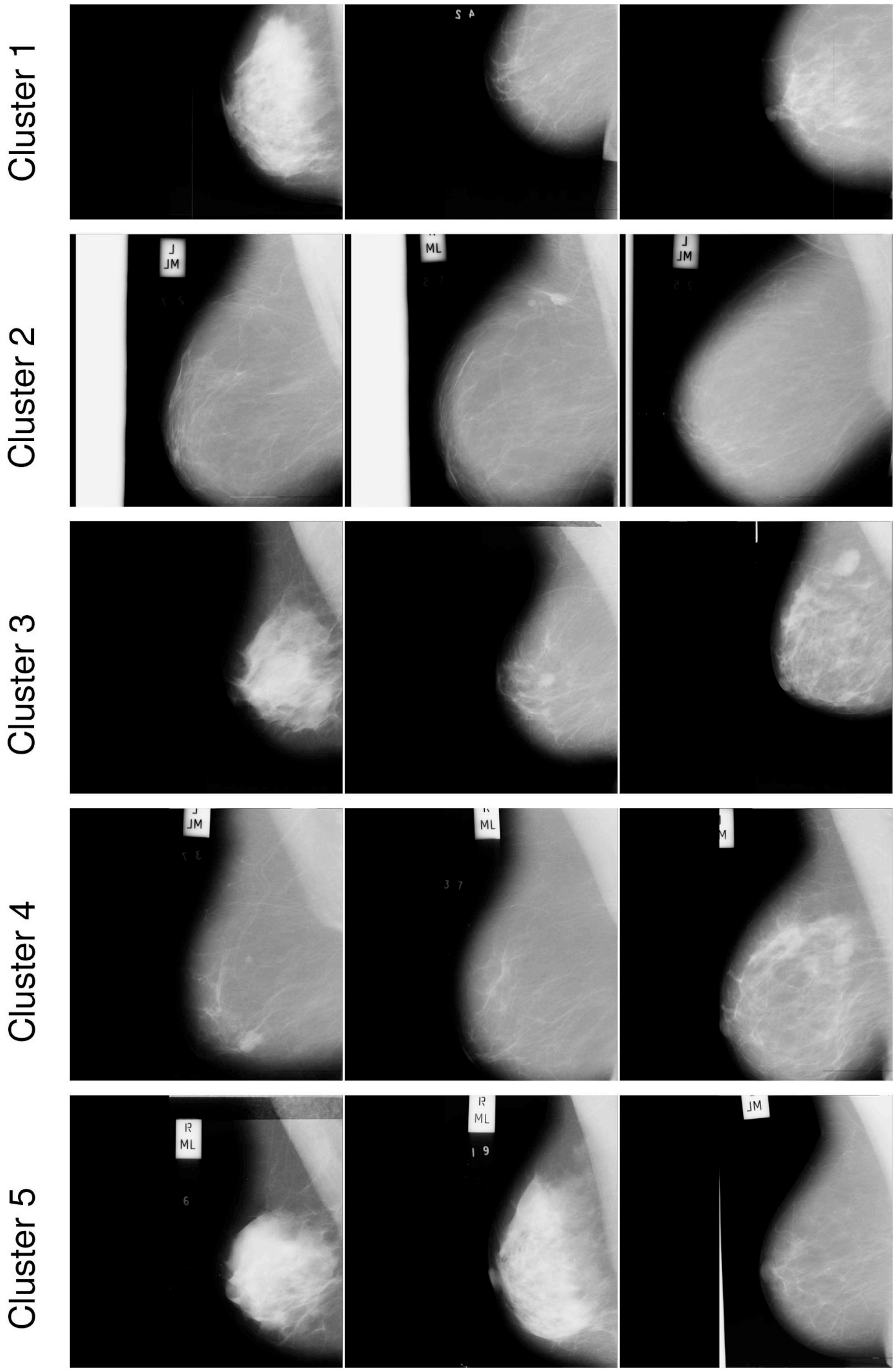


Fig. 3. Mammograms belonging to five different clusters of the projection in Fig. 2a. While images within a cluster are similar, images in different clusters have significantly different intensity profile and shape.

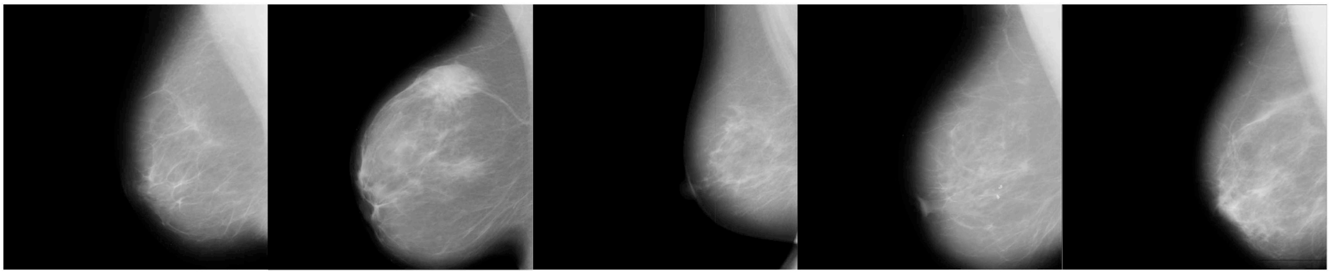


Fig. 4. The five atlases selected for each of the clusters in Fig. 3. Each atlas represents a different breast shape.

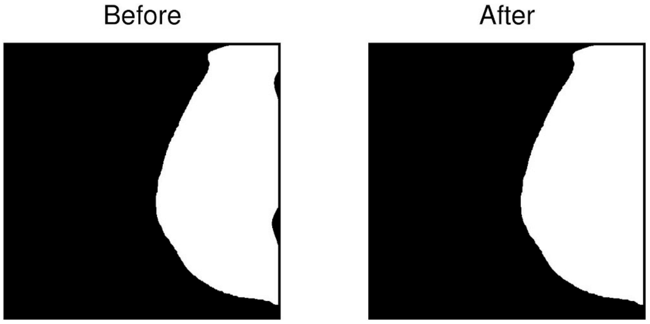


Fig. 5. Post-processing step to include breast regions lost during registration process. Here, we show the mask before and after the postprocessing for a particular mammogram.

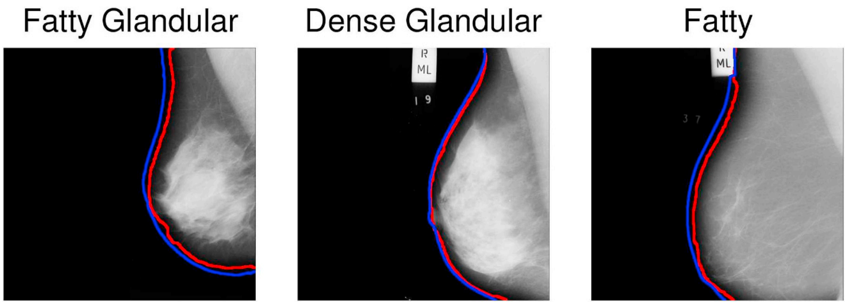
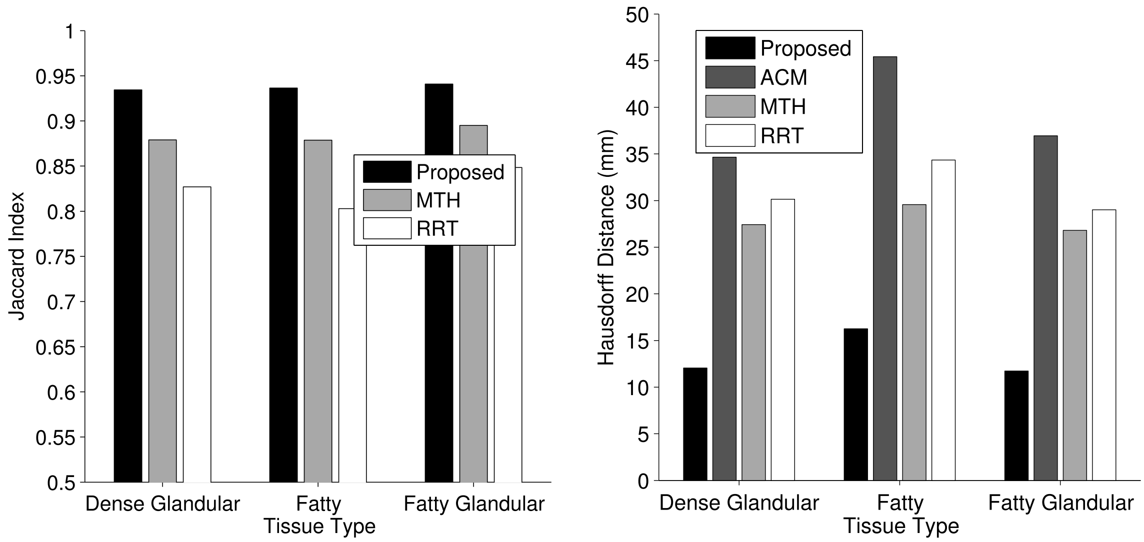


Fig. 6. Segmentation results for different tissue types in the mini-MIAS dataset. Each image represents one of the three types of tissues present. Blue boundary shows the ground truth and the red boundary shows the segmentation achieved. It can be seen that the proposed method is able to segment different types of breast tissues fairly close to the ground truth.



(a) Jaccard Index measures the degree of overlap between the ground truth and the computed segmentation. Higher is better.

(b) Average Hausdorff Distance. Lower is better.

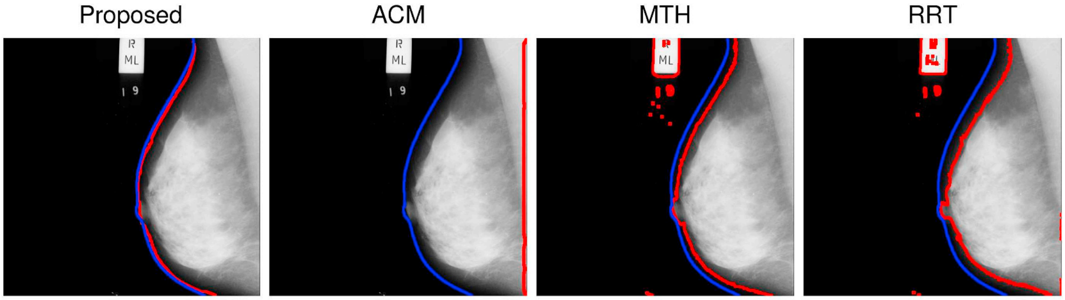
Fig. 7. Segmentation results computed on two different metrics for different tissue types in the mini-MIAS dataset. For ACM, we report only the Hausdroff distance as the ground truth masks are not available.

We used t-SNE [36] for identifying the number of clusters in the dataset. The t-SNE method preserves the neighborhood likelihood of two samples – similar points in the higher dimensional plane remain close in the projected dimensions and dissimilar points remain distant. However, the distances in the projected dimensions and the cluster sizes may not have any meaning [45]. Hence, clustering cannot be performed in the projected dimension. We performed clustering in the original image dimension of size $W \times H$. We verified that the clusters are preserved in Figs. 2a and 9a. We also verified in Fig. 3 that the mammogram images resemble each other within the clusters but are dissimilar between two clusters.

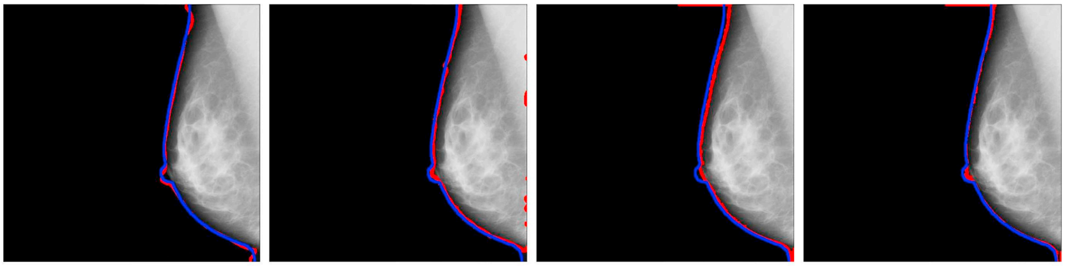
A potential limitation of our work is that the optimal number of clusters $k^\star$ depends on the t-SNE projection. Depending on the initialization of the t-SNE algorithm, the cluster shapes in the projection may differ [45]. The BIC plot, in that case, may yield a slightly different $k^\star$. One solution to this is to use a deterministic algorithm by either fixing the random seed or by using PCA projection for the initialization in 2D. Another alternative is to select the number of clusters $k^\star$ to be always slightly larger than the estimated minimum. Certainly, more atlas images can only be beneficial for the registration algorithm. This is also confirmed by BIC scores in Fig. 2b where the BIC for $k = 4$ is much larger than BIC for $k = 6$. Indeed, such heuristics are frequently used in the statistics literature in the context of model selection using cross-validation [46].

From Figs. 8 and 10, we notice that the thresholding-based methods are susceptible even to imperceptible salt and pepper noise near the breast boundary. On the other hand, the contour-based method ACM works better although it can sometimes be affected by contours due to external objects. Although segmentation based on deep learning is increasingly common [14–17], for it to be feasible on such a task, one would need a large number of segmented training images, even with pre-trained networks. In comparison, our method can handle salt and pepper noises, is less prone to be affected by external objects around the breast region and requires only a small number of segmented mammograms compared to the dataset size, 5–6 images only, *i.e.*approximately 1.5% for 322 images in the mini-MIAS dataset and about 1.1% for 532 images in the DDSM dataset.

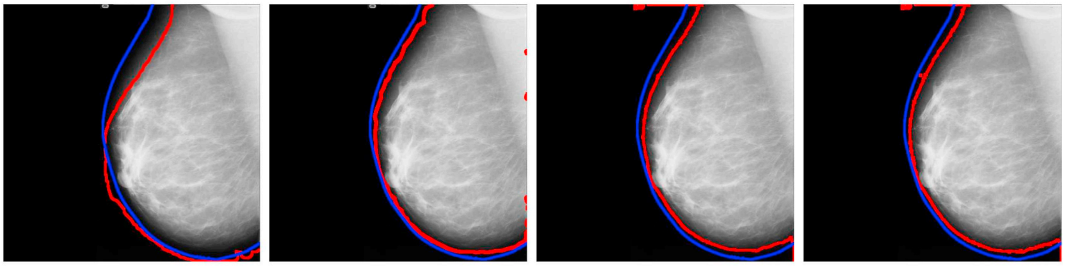
Finally, registration is computationally expensive, particularly for high-resolution images that are so common in mammography. We



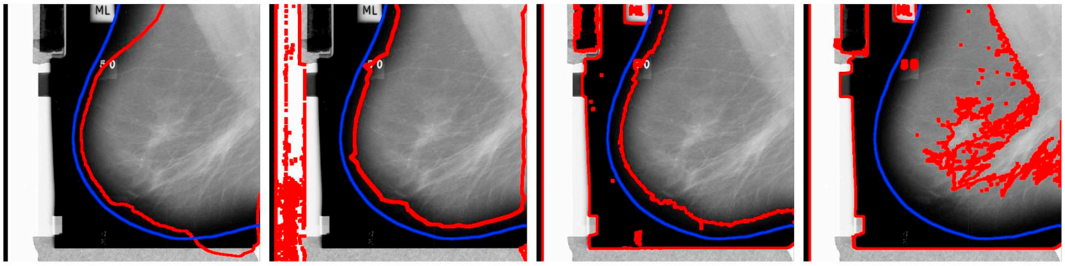
(a) Proposed method outperforms other methods



(b) All methods perform well



(c) ACM outperforms the proposed method

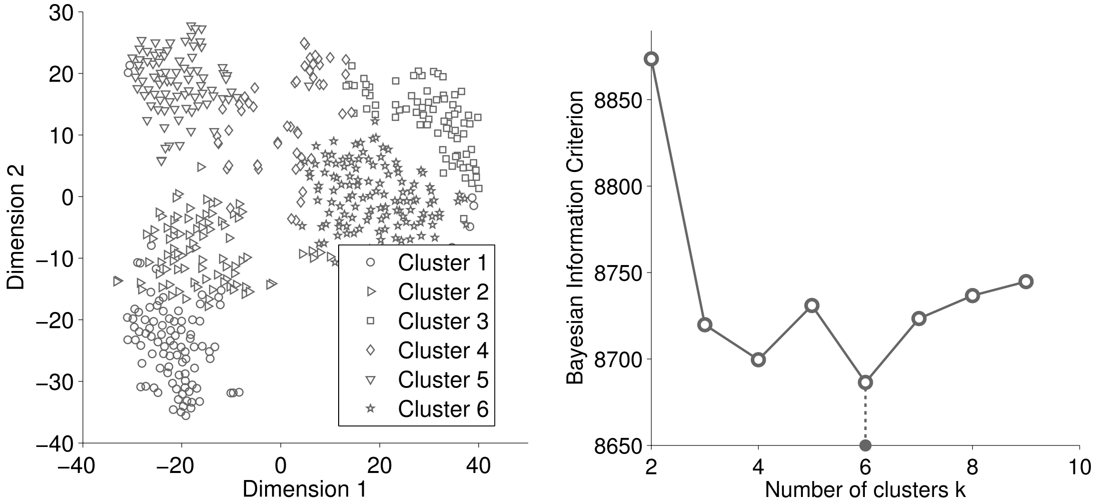


(d) None of the methods perform well

Fig. 8. Comparing breast region segmentation obtained using different methods. Methods in the order from left to right are, Proposed method, Active Contour Model (ACM), Multi-level thresholding (MTH) and Row-by-Row Thresholding (RRT). Here the red curve shows the segmentation while the blue curve shows the actual ground truth. We present four performance scenarios here. While each row show a different test image, images within a row show the segmentations obtained using different methods for the same test image.

Table 1
Comparison of segmentation results obtained with competing methods. Competing methods are Adaptive Contour Modelling (ACM), Multi-level Thresholding (MTH), Row-by-Row Thresholding (RRT).

Segmentation Method	Avg. Time (s)	Jaccard Index			Hausdorff Distance		
		Mean	Std. Deviation	Best in % of cases	Mean (mm)	Std. Deviation (mm)	Best in % of cases
Proposed Method	70.06	0.94	0.04	79.50	13.34	12.79	74.22
ACM [20]	18.69	–	–	–	38.94	14.45	3.42
MTH [2]	0.09	0.88	0.05	8.70	27.93	14.50	15.84
RRT [3]	0.93	0.83	0.14	11.80	31.16	15.86	6.52



(a) t-SNE projection of mammogram images to 2 dimensions.

(b) Bayesian Information Criterion (BIC) of GMMs vs. k, the number of Gaussians

Fig. 9. 2-D projection of images from the DDSM dataset and corresponding BIC curve. The minima of the BIC curve suggests the presence of $k^* = 6$ clusters. The distribution of six clusters is shown in the projected 2-D space.

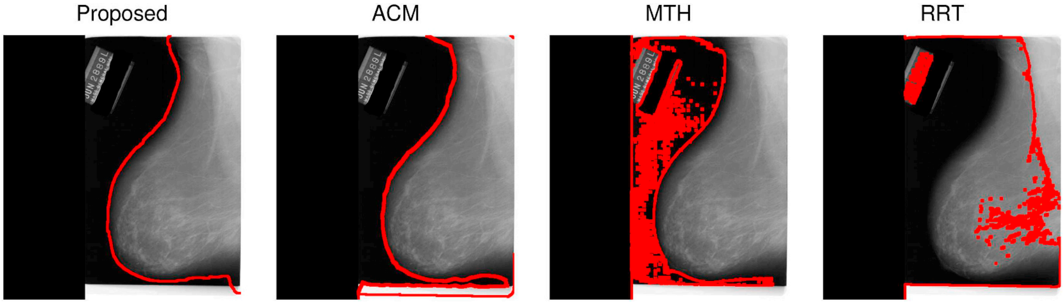


Fig. 10. Segmentation of breast region with different methods on an image from the DDSM dataset.

performed registration on 512 × 512 images and the time taken for registration is approximately 70 s per image which although greater when compared to other methods, allows to achieve better segmentation accuracy. To speed up the registration process, one can consider a more efficient registration strategy such as a multi-resolution approach [39] where the deformable registration is performed sequentially from a lower to a higher resolution or use a higher value of s_k in Equation (6) but the segmentation quality may be reduced in that case.

5. Conclusion

In this paper, we propose an atlas-based approach for segmenting breast region in mammograms. Due to the absence of clear landmarks, the variation in breast shape and intensity profiles, as well as scanner-related differences, it is hard to develop a segmentation method which generalizes well. This is particularly the case for many existing algorithms which consider each image independently of the others and thus cannot capture patterns in the dataset. Our approach is unique as we advocate that the atlas images must be selected carefully. Yet, as opposed to machine learning and especially deep learning methods, we do not need a large training dataset of segmented images. Furthermore, deformable image registration naturally accounts for the spatial structure of the image and not just the intensity profile when segmenting. Thus, our method, while outperforming three other methods on public datasets, is likely to work also on newer datasets. We demonstrate our results on two public datasets and two different comparison metrics. Finally, to facilitate the development of more advanced methods and for the benefit of the research community, the source code and segmented masks are publicly available[1].

Conflicts of interest

None declared.

References

[1] N. Otsu, A threshold selection method from gray-level histograms, Automatica 11 (285–296) (1975) 23–27.
[2] A.Z. Arifin, A. Asano, Image segmentation by histogram thresholding using hierarchical cluster analysis, Pattern Recogn. Lett. 27 (13) (2006) 1515–1521.
[3] A.A. AbuBaker, R. Qahwaji, M.J. Aqel, M.H. Saleh, Average row thresholding method for mammogram segmentation, Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the, IEEE, 2006, pp. 3288–3291.
[4] U. Bick, M.L. Giger, R.A. Schmidt, R.M. Nishikawa, D.E. Wolverton, K. Doi, Automated segmentation of digitized mammograms, Acad. Radiol. 2 (1) (1995) 1–9.
[5] D. Raba, A. Oliver, J. Martí, M. Peracaula, J. Espunya, Breast segmentation with pectoral muscle suppression on digital mammograms, Iberian Conference on Pattern Recognition and Image Analysis, Springer, 2005, pp. 471–478.
[6] M. Mustra, J. Bozek, M. Grgic, Breast border extraction and pectoral muscle detection using wavelet decomposition, IEEE EUROCON 2009, IEEE, 2009, pp. 1426–1433.
[7] J. Nagi, S.A. Kareem, F. Nagi, S.K. Ahmed, Automated breast profile segmentation for roi detection using digital mammograms, IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), IEEE, 2010, 2010, pp. 87–92.
[8] M. Mustra, M. Grgic, Robust automatic breast and pectoral muscle segmentation from scanned mammograms, Signal Process. 93 (10) (2013) 2817–2827.
[9] P. Casti, A. Mencattini, M. Salmeri, A. Ancona, F. Mangeri, M.L. Pepe, R.M. Rangayyan, Estimation of the breast skin-line in mammograms using multidirectional gabor filters, Comput. Biol. Med. 43 (11) (2013) 1870–1881.
[10] R.D. Yapa, K. Harada, Breast skin-line estimation and breast segmentation in mammograms using fast-marching method, Int. J. Biol. Biomed. Med. Sci. 3 (1)

(2008) 54–62.
[11] H.E. Rickard, G.D. Tourassi, N. Eltonsy, A.S. Elmaghraby, Breast segmentation in screening mammograms using multiscale analysis and self-organizing maps, Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, 1 IEEE, 2004, pp. 1786–1789.
[12] M. Wirth, J. Lyon, D. Nikitenko, A fuzzy approach to segmenting the breast region in mammograms, Fuzzy Information, 2004. Processing NAFIPS'04. IEEE Annual Meeting of the, 1 IEEE, 2004, pp. 474–479.
[13] R. Lajili, K. Kalti, A. Touil, B. Solaiman, N.E.B. Amara, Two-step evidential fusion approach for accurate breast region segmentation in mammograms, IET Image Process. 12 (11) (2018) 1972–1982.
[14] G. Carneiro, J. Nascimento, A.P. Bradley, Unregistered multiview mammogram analysis with pre-trained deep learning models, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 652–660.
[15] D. Lévy, A. Jain, Breast Mass Classification from Mammograms Using Deep Convolutional Neural Networks, arXiv preprint arXiv:1612.00542.
[16] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, L. Li, Discrimination of breast cancer with microcalcifications on mammography by deep learning, Sci. Rep. 6 (2016) 27327.
[17] K. J. Geras, S. Wolfson, S. Kim, L. Moy, K. Cho, High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks, arXiv preprint arXiv:1703.07047.
[18] M. Jas, J. Chakrobarty, S. Mukhopadhyay, An atlas based approach to segmenting MLO view mammograms, Informatics, Electronics & Vision (ICIEV), 2012 International Conference on, IEEE, 2012, pp. 7–11.
[19] R. Martí, A. Oliver, D. Raba, J. Freixenet, Breast skin-line segmentation using contour growing, Iberian Conference on Pattern Recognition and Image Analysis, Springer, 2007, pp. 564–571.
[20] R. Ferrari, A. Frere, R. Rangayyan, J. Desautels, R. Borges, Identification of the breast boundary in mammograms using active contour models, Med. Biol. Eng. Comput. 42 (2) (2004) 201–208.
[21] A. Rampun, P.J. Morrow, B.W. Scotney, J. Winder, Fully automated breast boundary and pectoral muscle segmentation in mammograms, Artif. Intell. Med. 79 (2017) 28–41.
[22] M.A. Wirth, A. Stapinski, Segmentation of the breast region in mammograms using snakes, Computer and Robot Vision, 2004. Proceedings. First Canadian Conference on, IEEE, 2004, pp. 385–392.
[23] K.L. Kashyap, M.K. Bajpai, P. Khanna, G. Giakos, Mesh-free based variational level set evolution for breast region segmentation and abnormality detection using mammograms, Int. J. Numer. Methods Biomed. Eng. 34 (1) (2018) e2907.
[24] M. Karnan, K. Thangavel, Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications, Comput. Methods Progr. Biomed. 87 (1) (2007) 12–20.
[25] K. Wei, W. Guangzhi, D. Hui, Segmentation of the breast region in mammograms using watershed transformation, Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the, IEEE, 2006, pp. 6500–6503.
[26] P. Shi, J. Zhong, A. Rampun, H. Wang, A hierarchical pipeline for breast boundary segmentation and calcification detection in mammograms, Comput. Biol. Med. 96 (2018) 178–188.
[27] N. Saidin, U.K. Ngah, H.A.M. Sakim, D.N. Siong, M.K. Hoe, I.L. Shuaib, Density based breast segmentation for mammograms using graph cut and seed based region growing techniques, Second International Conference on Computer Research and Development, IEEE, 2010, 2010, pp. 246–250.
[28] M. Mustra, M. Grgic, R.M. Rangayyan, Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms, Med. Biol. Eng. Comput. 54 (7) (2016) 1003–1024.
[29] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, M.B. Cuadra, A review of atlas-based segmentation for magnetic resonance brain images, Comput. Methods Progr. Biomed. 104 (3) (2011) e158–e177.
[30] I. Isgum, M. Staring, A. Rutten, M. Prokop, M.A. Viergever, B. Van Ginneken, Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans, IEEE Trans. Med. Imaging 28 (7) (2009) 1000–1010.
[31] L. Zhang, E. Hoffman, J.M. Reinhardt, Atlas-driven lung lobe segmentation in volumetric X-ray CT images, Medical Imaging, IEEE Trans. 25 (1) (2006) 1–16.
[32] M. Jas, S. Mukhopadhyay, J. Chakraborty, A. Sadhu, N. Khandelwal, A heuristic approach to automated nipple detection in digital mammograms, J. Digit. Imaging 26 (5) (2013) 932–940.
[33] J. Chakraborty, S. Mukhopadhyay, V. Singla, N. Khandelwal, P. Bhattacharyya, Automatic detection of pectoral muscle using average gradient and shape based feature, J. Digit. Imaging 25 (3) (2012) 387–399.
[34] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, The mammographic image analysis society digital mammogram database, Exerpta Medica. International Congress Series, 1069 1994, pp. 375–378.
[35] M. Wu, C. Rosano, P. Lopez-Garcia, C.S. Carter, H.J. Aizenstein, Optimum template selection for atlas-based segmentation, Neuroimage 34 (4) (2007) 1612–1618.
[36] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2579–2605) (2008) 85.
[37] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (2) (1978) 461–464.
[38] J.-P. Thirion, Image matching as a diffusion process: an analogy with Maxwell's demons, Med. Image Anal. 2 (3) (1998) 243–260.
[39] H. Wang, L. Dong, J. O'Daniel, R. Mohan, A.S. Garden, K.K. Ang, D.A. Kuban, M. Bonnen, J.Y. Chang, R. Cheung, Validation of an accelerated'demons' algorithm for deformable image registration in radiation therapy, Phys. Med. Biol. 50 (12) (2005) 2887.
[40] S. Saeb, L. Lonini, A. Jayaraman, D.C. Mohr, K.P. Kording, The need to approximate the use-case in clinical machine learning, GigaScience 6 (5) (2017) 1–9.
[41] X. Pennec, P. Cachier, N. Ayache, Understanding the ”demon's algorithm”: 3d non-rigid registration by gradient descent, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 1999, pp. 597–605.
[42] P. Jaccard, Etude comparative de la distribution florale dans une portion des Alpes et du Jura, Impr. Corbaz, 1901.
[43] F. Hausdorf, J.R. Aumann, Grundzüge der mengenlehre, Veit, (1914).
[44] M. Wirth, J. Lyon, M. Fraschini, D. Nikitenko, The effect of mammogram databases on algorithm performance, Null, IEEE, 2004, p. 15.
[45] M. Wattenberg, F. Viégas, I. Johnson, How to use t-sne effectively, Distill 1 (10) (2016) e2.
[46] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw. 33 (1) (2010) 1.