Multi-scale classification based lesion segmentation for dermoscopic images

Mani Abedini, Noel Codella, Rajib Chakravorty, Rahil Garnavi, *Member, IEEE,*
David Gutman, Brian Helba, and John R. Smith,

***Abstract*— This paper presents a robust segmentation method based on multi-scale classification to identify the lesion boundary in dermoscopic images. Our proposed method leverages a collection of classifiers which are trained at various resolutions to categorize each pixel as "lesion" or "surrounding skin". In detection phase, trained classifiers are applied on new images. The classifier outputs are fused at pixel level to build probability maps which represent lesion saliency maps. In the next step, Otsu thresholding is applied to convert the saliency maps to binary masks, which determine the border of the lesions. We compared our proposed method with existing lesion segmentation methods proposed in the literature using two dermoscopy data sets (International Skin Imaging Collaboration and Pedro Hispano Hospital) which demonstrates the superiority of our method with Dice Coefficient of 0.91 and accuracy of 94%.**

I. Introduction

Malignant melanoma is one of the most common and the deadliest types of skin cancer [1], [2]; it has been shown that early detection of melanoma and immediate surgical excision of the lesion can increase the survival rate by five fold [3]. Dermoscopy is an important non-invasive tool for clinical diagnosis of melanoma. However, due to subjectivity of human expert in the interpretation of skin lesions, computer-based analysis of dermoscopic images has become more popular in the last decade. Skin lesion border detection is the first and a very crucial step towards automated dermoscopic image analysis.

Border detection in dermoscopic images is a challenging task, and a lot of effort has been made to develop accurate lesion segmentation methods. Generally, lesion segmentations are categorized into three groups: thresholding, edge based and region based methods.

Silveira et al [4] present a comprehensive survey on existing methods developed for lesion segmentation, including Gradient Vector Flow (GVF) [5], the Level Set method of Chan et al. [6], Adaptive Thresholding (AT) [4], Adaptive Snakes (AS) [7], Expectation Maximization Level Set [8], and Fuzzy-based Split and-Merge algorithm [9].

Thresholding methods achieve good results when there is good contrast between the lesion and surrounding skin. AT is one of the well-known thresholding methods. Ganster et.al proposed a fusion of global thresholding, AT and clustering method in [10]. GVF and AS are among popular edge-based segmentation methods. In [5] an edged based segmentation methods based on zero-crossing of Laplacian-of-gaussian and geodesic active contour has been applied for lesion segmentation. From region based methods, Level-set Active Contours (LSAC) [11] and Statistical Region Merging (SRM) [12] are proposed and frequently referenced in the literature. LSAC is a state-of-the-art region-based active contour segmentation algorithm which has been used for lesion segmentation in [13].

In this paper, we propose a multi-scale classification based border detection method to identify the lesion boundary and separate the lesion from surrounding skin. The reason of proposing multi-scale classification process is that we noted that spatial information and the structure of neighbouring pixels can contribute towards more accurate lesion segmentation. Thus, our proposed technique leverage having a very top level map of the skin and lesion (lesion localization) as well as very detail predictions for pixels along the border. The intermediate levels can help constructing the final saliency map.

In order to validate our method, we have compared it with state-of-the art techniques reported in [14] using Pedro Hispano Hospital dataset (PH^2) [15] as benchmark. Additional experimental study has been conducted to compare the proposed method with three widely used border segmentation methods; namely, AT, SRM, and LSAC (one method from each of above-mentioned categories). We have also implemented three conventional thresholding based object segmentation methods to ascertain the accuracy of our method. Moreover, to assess the robustness of our method, we have tested the performance on an a second dataset obtained through collaboration with the International Skin Imaging Collaboration (ISIC) [16].

The rest of the paper is organized as follows: Section II explains the proposed method, Section III discusses the experimental methodology and achieved results. Section IV concludes the paper.

II. Methods

Our proposed approach is comprised of two phases:

Training phase: First, all training images (in RGB color channel) are broken down into several patches using Simple Linear Iterative Clustering (SLIC) super pixel generator (The number of patches are defined by λ). Next, the label of each patch is determined (The majority pixel label of the patch). Then, the training set is divided into positive and negative sets. Next, low-level features of the training set are extracted

M. Abedini, R. Chakravorty and R. Garnavi are with IBM Research - Australia, Melbourne, VIC, Australia, e-mails: [mabedini,rachakra,rahilgar]@au1.ibm.com.

N. Codella and J. R. Smith are with IBM T.J. Watson Research Center, Yorktown Heights, NY, USA, e-mails: [nccodell,jsmith]@us.ibm.com.

D. Gutman is with Emory University School of Medicine, Atlanta, GA, USA, e-mail: dgutman@emory.edu.

B. Helba is with Kitware, USA, e-mail: brian.helba@kitware.com.

978-1-4577-0220-4/16/$31.00 ©2016 IEEE

(e.g. color correlogram, wavelet texture, color histogram, color wavelet, edge histogram, GIST, multi-scale local binary pattern). Finally, an ensemble based classifier is trained using IBM Multimedia Analysis and Retrieval System (IMARS) machine learning tool on the generated features. The system has a bagging mechanism to split training data into bags and train several SVM classifiers. The system has internal feature selection and final model fusion mechanism to avoid the curse of dimensionality and improve the accuracy of the trained models. For further detail about the IMARS tool we refer readers to [17], [18]. This process is repeated in different scales (λ) which will be resulted in constructing multi-scale classifiers (see Figure 1).

One of the advantages of this method is to discard noise objects (such as hair/rulers) as well as unwanted artifacts (such as fiducial markers) during recognition phase. In other words, RGB image is directly given to the border detection method without any further preprocessing steps, such as hair removal or fiducial marker removal.

During learning phase, the classifiers are trained with positive and negative samples with and without the presence of noise objects (e.g. hair/rulers/fiducial markers). Thus, the classifiers learn patterns/features which are discriminative for mole patches from skin patches and ignore common patterns such as hair/ rulers. Also, big objects such as fiducial markers are discarded automatically because inherent properties of mole patches are very different from properties of artificial artifacts such as fiducial markers.

Detection phase: First, a given image is decomposed to λ patches by using SLIC super pixel algorithm. Then, the trained classifier associated with λ are applied on the extracted patches. By putting back the predicted label of each patch, a prediction label-map is constructed. By repeating the same process for various λs, several prediction label-maps are generated. In order to merge the generated prediction label-maps, we calculate the average of pixel labels; which results in a combined probability map. The map represent saliency map of the lesion. Then, the generated map is thresholded to generate a binary mole border map (see Figure 2). Figure 4 shows a sample dermoscopic image, the generated border by our method and its ground truth delineated by dermatologists.

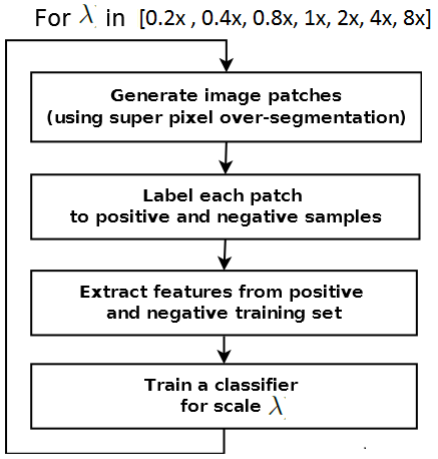
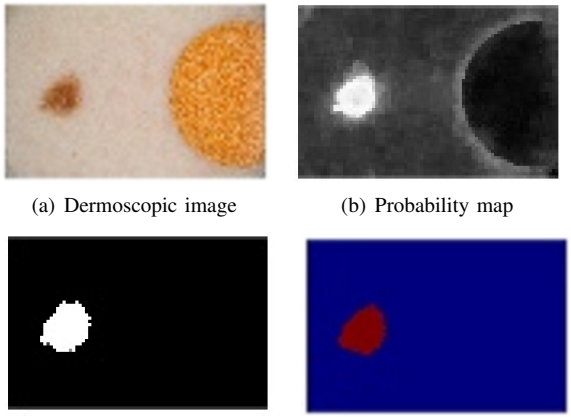


Fig. 1. In training phase, for various λ, our proposed method breaks down the image into super pixels of size λ then train a classifier.



(a) Dermoscopic image (b) Probability map

(c) Generated border (d) Ground truth border

Fig. 4. Dermoscopic image (a), the saliency map as fusion of classifier outputs at pixel level (b), the binary mark after applying thresholding on the saliency map (c), the ground truth border (d).

III. Experiment Methodology and Results

Dermatology datasets used for evaluation of our method were on PH^2 (with 200 images) and ISIC data set (with 500 images).

For each set, we run the proposed method on 2 fold-cross validation technique for various λ= (0.2x, 0.4x, 0.8x, 1x, 2x, 4x, 8x) where (1x = 100 patches). The generated borders of each level as well as the final merged border have been compared with ground truth (generated by domain experts) in terms of Average Precision (AP), Accuracy (ACC), Jaccard index (JC), and Dice coefficient (DC).

Our proposed method has been evaluated on PH^2, then compared with published results in [14], as well as with conventional lesion segmentation methods. The numbers suggest that our method outperformes state-of-the-art border detection methods reported on PH^2 benchmark. Since, ISIC data set has not been used in any similar study at the time of publishing this paper, we run comparative study between our method and four conventional thresholding based border detection methods. Also, in order to investigate the comparative performance of the thresholding based border detection methods as the baseline methods; we run the same experiments on PH^2. The results are reported in Tables I and II.

The thresholding methods has a straightforward pipeline; First images are cleaned after hair removal process. Then, the blue channel is passed to thresholding method (Otsu, Adaptive Thresholding, ISO Data, Yen thresholding). It was experimentally found that the blue component in the RGB representation allows more discrimination of boundary between lesion and skin in most dermoscopic images. Finally,

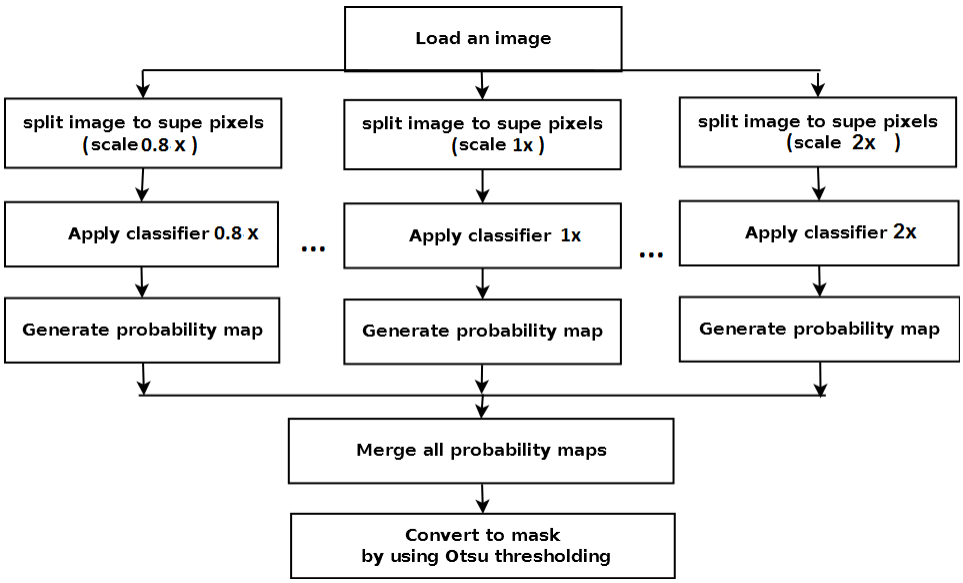


Fig. 2. In detection phase, for various λ, out proposed algorithm breaks down the test image into small super pixels and run associated classifier to generate probability maps. After merging all generated probability maps, Otsu thresholding is applied to build a binary mask. The binary mask represents the lesion and its border.

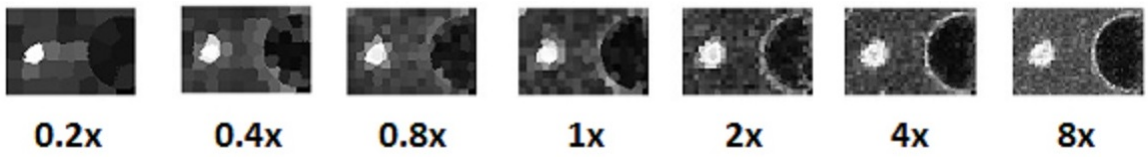


Fig. 3. For various λ values, a prediction label map is generated. In our experiments, the initial λ is 100, which means each dermoscopic image is broken down to 100 patches. λ is then scaled up (multiplying factors of 2, 4, 8) and down (multiply factors of 0.4, 0.8, 0.2).

as the post processing step only objects that are most likely to be moles are kept and other objects are discarded. To do that, the image (in RGB color) is divided into 20 patches and examined by a mole classifier (similar classifier as trained in Section II. The region that has the highest probability of being mole is identified and all objects that have no intersection with this region are discarded. We run the same process to find the mole localization and pick one seeding point inside the lesion to run AT, SRM, and LSAC. In our experiments, we noted that our automated seeding point process has 99% accuracy. This process is necessary because some algorithms needs intial seeding point (such as LSAC), or to assist other algorithm to avoid noise objects such as fiducial markers by defining a bounding box around lesion. The ISIC data set contains more challenging due to the presence of hair and noise artifacts in the images.

By comparing the results, we can see that our implement thresholding based border detection methods have comparative performance with state of the art methods reported in the literature. Comparative results between our proposed method and reported results in [14] on PH^2 dataset suggest that our proposed method is performing better than all other methods in terms of Dice metric as follow: Adaptive thresholding= 0.80, Chan-based level set= 0.71, Region growing border detection=0.61, Saliency based border detection=0.84 and Our proposed method 0.91.

We used this conclusion to interpret Table II. The drop off in the performance of all border detection methods suggest that ISIC data set is relatively harder than PH^2. One reason is the presence of hair and fiducial markers as well as the fact that most images in ISIC data have low contrast between mole lesion and surrounding skin.

The results of experiments on ISIC data (Table II) suggest that our border detection method has better accuracy compared to other methods. We also investigated if the fusion of the thresholding methods and our proposed segmentation method can improve the accuracy further. The evaluation of borders generated from the weighted fusion of all methods demonstrates that the fusion process does not improve the accuracy of generated borders (See last row of Tables I and II).

IV. Conclusion

In this paper, we presented a robust and accurate method for detecting lesion border in dermoscopic images. Our comparative study conducted on two different datasets demonstrated that our proposed method outperform state-of-the art lesion segmentation techniques as well as conventional thresholding based border detection methods. With respect to associated costs; our method is incorporating supervised learning mechanism which demands time and computation resources for training. Further study on fusion of all results is

TABLE I

THE COMPARISON RESULTS BETWEEN OUR PROPOSED METHOD AND OTHER BORDER DETECTION METHODS ON PH^2. THE HIGHEST SCORES OF EACH METRIC IS MARKED IN BOLDFACE.

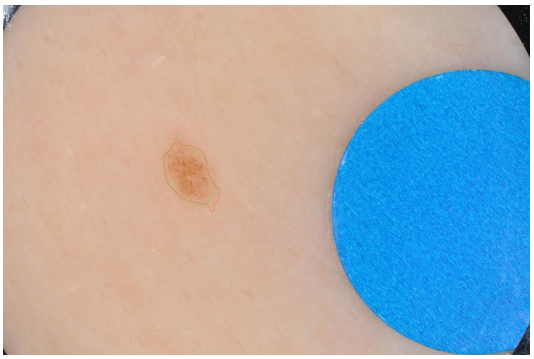
Methods	AP	ACC	JC	DC
Otsu	0.89	0.89	0.74	0.84
ISO Data Thresholding	0.89	0.89	0.73	0.83
Yen Thresholding	0.87	0.84	0.70	0.79
Adaptive Thresholding [4]	0.87	0.89	0.72	0.80
Level set Active Contours [11]	0.86	0.85	0.76	0.83
Statistical Region Merging [12]	0.89	0.92	0.78	0.86
Our proposed method	**0.93**	**0.94**	**0.84**	**0.91**
Fusion of all	0.92	0.91	0.80	0.88

TABLE II

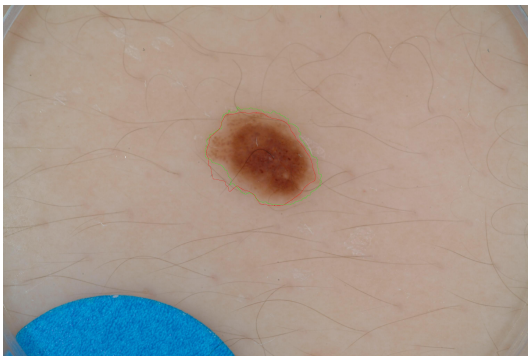
THE COMPARISON RESULTS BETWEEN OUR PROPOSED METHOD AND OTHER BORDER DETECTION METHODS ON ISIC DATASETS. THE HIGHEST SCORES OF EACH METRIC IS MARKED IN BOLDFACE.

Methods	AP	ACC	JC	DC
Otsu	0.70	0.96	0.50	0.60
ISO Data Thresholding	0.62	0.92	0.33	0.44
Yen Thresholding	0.67	0.93	0.45	0.55
Adaptive Thresholding	0.72	0.98	0.45	0.59
Level set Active Contours	0.70	0.95	0.43	0.53
Statistical Region Merging	0.74	0.98	0.46	0.60
Our proposed method	**0.79**	**0.98**	**0.61**	**0.72**
Fusion of all	0.73	0.97	0.55	0.67

being investigated which demonstrated that the overall performance of lesion segmentation has been improved. However, the accuracy of the multi-scale classification method is marginally better than the fusion model. Thus, we conclude that the multi-scale classification based border detection method is robust enough for segmenting lesion area from the surrounding skin. We also observed that our method is resilient to present noise objects in images (e.g. hair, fiducial markers, etc).



(a) Example 1



(b) Example 2

Fig. 5. Two examples of borders generated by our proposed border detection method (presented by red contour). The green contour in the images represents the ground truth drawn by dermatologists.

REFERENCES

[1] Cancer Facts & Figures 2014. American Cancer Society, 2014.

[2] R. Braun, H. Rabinovitz, M. Oliviero, A. Kopf, and J. Saurat. Dermoscopic of pigmented lesions. Journal of the American Academy of Dermatology, vol. 52, no. 1, pp. 109121, 2005.

[3] A. C. Geller, S. M. Swetter, K. Brooks, M.-F. Demierre, and A. L. Yaroch, Screening, early detection, and trends for melanoma: Current status (2000-2006) and future directions, Journal of the American Academy of Dermatology, vol. 57, pp. (555-572), 2007.

[4] SilveiraM. and Nascimento, J.C. and Marques, J.S. and Marcal, A.R.S. and Mendonca, T. and Yamauchi, S. and Maeda, J. and Rozeira, J., "Comparison of Segmentation Methods for Melanoma Diagnosis in Dermoscopy Images", IEEE Journal of Selected Topics in Signal Processing, Vol. 3, number 1, pp 35-45, 2009.

[5] Erkol B, Moss RH, Stanley RJ, Stoecker WV, Hvatum E, "Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes", Skin Res Technol. 11(1): pp 17-26, 2005.

[6] T. Chan, B. Sandberg, and L. Vese, "Active contours without edges for vector-valued images," J. Vis. Commun. Image Repres., vol. 11, no. 2, pp. 130141, 2000.

[7] Nascimento JC, Marques JS, "Adaptive snakes using the EM algorithm", IEEE Trans Image Process. 14(11):1678-86, 2005.

[8] G. McLachlan and T. Krishnan, "The EM Algorithm and Extensions." New York: Wiley, 1997.

[9] J. Maeda, A. Kawano, S. Saga, and Y. Suzuki, "Unsupervised perceptual segmentation of natural color images using fuzzy-based hierarchical algorithm," in Proc. SCIA. New York: Springer, 2007, vol. 4522, Lecture Notes in Computer Science, pp. 462471.

[10] Ganster H, Pinz A, Rhrer R, Wildling E, Binder M, Kittler H., "Automated melanoma recognition", IEEE Trans Med Imaging, 20(3):pp 233-9, 2001.

[11] Li C, Kao CY, Gore JC, Ding Z,"Minimization of region-scalable fitting energy for image segmentation." IEEE Trans Image Process. 17(10):1940-9, 2008.

[12] M. E. Celebi, H. A. Kingravi, H. Iyatomi, A. Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz, and S. W. Menzies, "Border Detection in Dermoscopy Images Using Statistical Region Merging," Skin Research and Technology, vol. 14, no. 3, pp. 347-353, 2008.

[13] Wong A, Scharcanski J, Fieguth P, "Automatic skin lesion segmentation via iterative stochastic region merging." IEEE Trans Inf Technol Biomed. 15(6):929-36, 2011.

[14] Euijoon Ahn, Youn Hyun Jung, Dr. Jinman Kim. Automated Saliency-based Melanoma Detection in Dermoscopic Images, Research Conversazione 2014, The university of Sydney, 2014.

[15] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. Marcal, J. Rozeira, PH2 - a dermoscopic image database for research and benchmarking. IEEE Conference Engineering in Medicine and Biology Society, pp. (5437-5440), 2013.

[16] International Skin Imaging Collaboration Website. Available: http://www.isdis.net/index.php/isic-project

[17] N. C. F. Codella, A. Natsev, G. Hua, M. Hill, L. Cao, L. Gong, and J. R. Smith, "Video event detection using temporal pyramids of visual semantics with kernel optimization and model subspace boosting," in Multimedia and Expo (ICME), 2012 IEEE International Conference on, pp. 747752, IEEE, 2012.

[18] Mani Abedini, Qiang Chen, Noel Codella, Rahil Garnavi, Xingzhi Sun, "Accurate and Scalable System for Automatic Detection of Malignant Melanoma", Dermoscopy Image Analysis, Ed: M. Emre Celebi, Teresa Mendonca, Jorge S. Marques, CRC Press , 2015.