Skin Cancer Segmentation

Yipeng Zhang,
Electrical and Computer Engineering
University of California, San Diego
California, CA 92037
yiz108@ucsd.edu

Abstract

In this paper, the project of Skin cancer segmentation is demonstrated which I explored during this quarter under the supervision of Professor Vwani Roychowdhury from University of California, Los Angeles. In this project, after the problem is stated, the review of the objective and the datasets are used is introduced. The methodology explains the procedure of the logic pipeline and the result and analysis section gives a comprehensive explanation of the dataset along with the performance of the algorithm.

1 Introduction

The task of this project is to analyze melanoma pictures, and then segment the melanoma images to get the parts corresponding to the disease. A further splitting of the masked image is perform to generate different shapes of the masks, The results will be further used by another research group to analyze the relationship between shapes and disease. The final conclusion of the project is the algorithm can successfully segment most of the images in the dataset, and even out-performs the segmentation labels. However, there is still room for the algorithm to be more optimized.

2 Objective

The objection of this project is, given a dermatoscopic images, extracting skin cancer area in the picture as accurate as possible, and generate masks on different intensity levels based on the segmentation result, so that the other group can use the generated mask to analyze the relationship between shapes and disease.

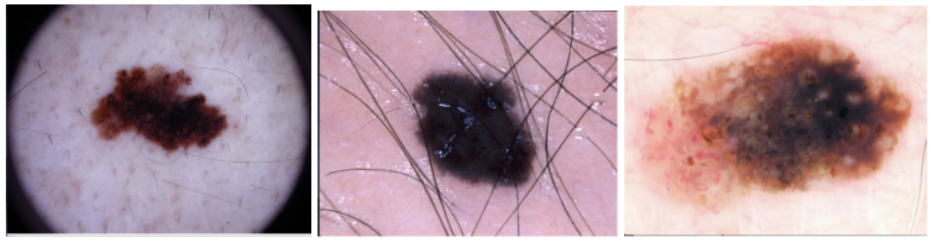
3 Dataset

There are total two datasets are used, HAM10000[1], and a dataset from ISIC 2018 Lesion Boundary Segmentation[2]. The HAM10000 dataset contains 10015 dermatoscopic images without labels of segmentation. ISIC 2018 Lesion Boundary Segmentation dataset contains 2594 dermatoscopic images with their corresponding segmentation results. According to the ISIC 2018 Lesion Boundary Segmentation challenge[2], all of the segmentation label are reviewed and curated by practicing dermatologists with expertise in dermoscopy so the labels can be used to justify the performance of the algorithm.

3.1 Dataset Visualization

Both datasets contain a decent amount of dermatoscopic images. Due to the fact that classic image processing approach is used to design the algorithm, the dermatoscopic images have

to be analyzed. Three examples of very noisy images shown in Figure 1 are presented to illustrate what the algorithm should cover to achieve successful segmentation.



(a) (b) (c)

Figure 1: The noise in the pictures

From Figure 1 above, it can be seen that there are the shadow of the camera(the black area surrounding the skin area.) in Figure 1 a, hairs in Figure 1 b and an unclear boundary in Figure 1c. These are three major takes I need to resolve in the image processing algorithm to perform a very successful segmentation.

However, the ISIC dataset contains more noise than HAM1000 dataset which has not only hair but some irrelevant objects for example, the disk with different color than the skin, which will interference the segmentation. The Figure 2 demonstrates the different noise in different dataset.

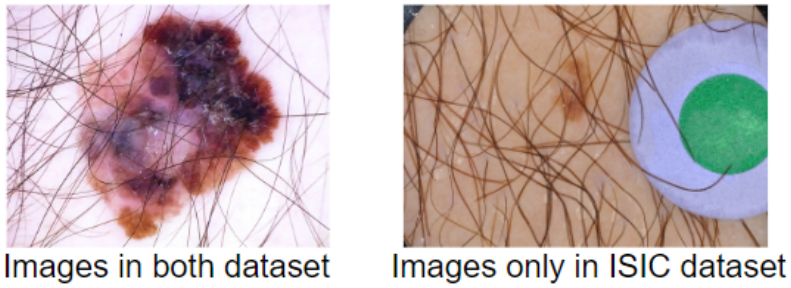


Figure 2: Image sample comparison in different datasets

4 Methodology

The method used in this project includes two parts, segmentation and mask extraction. The skin cancer segmentation includes image preprocessing, segmentation and result post processing, and the mask extraction includes a local histogram equalization which smooths the extraction procedure.

4.1 Skin Cancer Segmentation

4.1.1 Data-Preprocessing

Due to the fact that the noise in the image includes hair removal, scope shadow removal, the data preprocessing includes algorithm covering these two aspects.

Hair removal is using DullRazor algorithm [3]. First it uses the 3 structure elements(vertical, horizontal, and 45 degree) to locate the hair. Then the algorithm performs denoising and interpolation of the removed hair. More specifically, the algorithm first checks all hair pixels in the image to remove the outliers with opening and closing morphological operations. After

the hair is removed, the algorithm checks all surrounding pixels around the removed pixel to interpolate the best possible value for the removed pixel. Finally a median filter is performed to remove the artifacts generated from the procedure above.

Scope shadow removal is first detecting the area that connect to the border of the image using bounding box, then remember these area so that in the segmentation stage, these area does not take account.

4.1.2 Segmentation

The segmentation is using the Otsu method [4]. The Ostu method is used to automatically perform clustering-based image thresholding. Otsu is to segment the image into two clusters, and the intensity value histogram can be viewed as a bi-modal histogram (foreground pixels and background pixels).By performing Otsu method, it calculates the optimum threshold separating the two classes so that their intra-class variance is minimal. The Otsu method performs extremely well in segmenting images.

4.1.3 Segmentation Post Processing

The post processing includes three stage as well. It fist fills the small holes in the regions. Then it calculated area of each connected components, and removes the area with very small size. Then the area of the shadow of the scope won't be calculated since they are already detected as not the skin cancer.

4.2 Mask Extraction

The mask extraction is to split of the segmentation results within the mask to generate different shapes of the lesion based on their intensity value. The main objective of extracting mask it to make each level of extraction has the similar number of pixels. Therefore, a histogram equalization within the mask segmentation mask is needed, which only counts the value within the segmentation mask, and them smooths the histogram. After the histogram obey the general uniform distribution, threshold the histogram evenly into a predefined value. Finally, generates different level of mask based on the thresholds.

5 Results and Analysis

The original purpose of this project is generating different level of mask based on the HAM10000 dataset. However, it is very hard to define the performance of the algorithm because the area of skin cancer is very hard to be identify in some images. Therefore, the use of the ISIC 2018 Lesion Boundary Segmentation dataset along with the segmentation labels is to justify the performance of the algorithm.

5.1 Evaluation criteria

Since there are no evaluation criteria for mask generation, the main purpose of evaluation is to evaluate the segmentation result. There are total three criteria are used to justify the performance of the algorithm. First is the Jaccard index, which is defined as intersection over union of segmentation result and label, which is defined in the equation 1.

$$J(SegmentationResult, Label) = \frac{SegmentationResult \cap Label}{SegmentationResult \cup Label} \tag{1}$$

The intersection over segments(IOS) is also used in evaluating the performance, the equation is defined in the equation 2

$$IOS(SegmentationResult, Label) = \frac{SegmentationResult \cap Label}{SegmentationResult} \tag{2}$$

The last evaluation criteria is human inspection, which is also very important to evaluate difference between labels and segmentation results.

5.2 Good Result and Bad Results

Some good segmentation results and bad segmentation results are shown here, along with the discussion of the reason behind the results. In every result, the comparison between labels and segmentation results is shown in green circle and red circle respectively.

5.2.1 Good Result

Figure 3 displays the good results generated by the algorithm. Two evaluation criteria are both very high in these images, which means the algorithm segments images exactly same as the labels

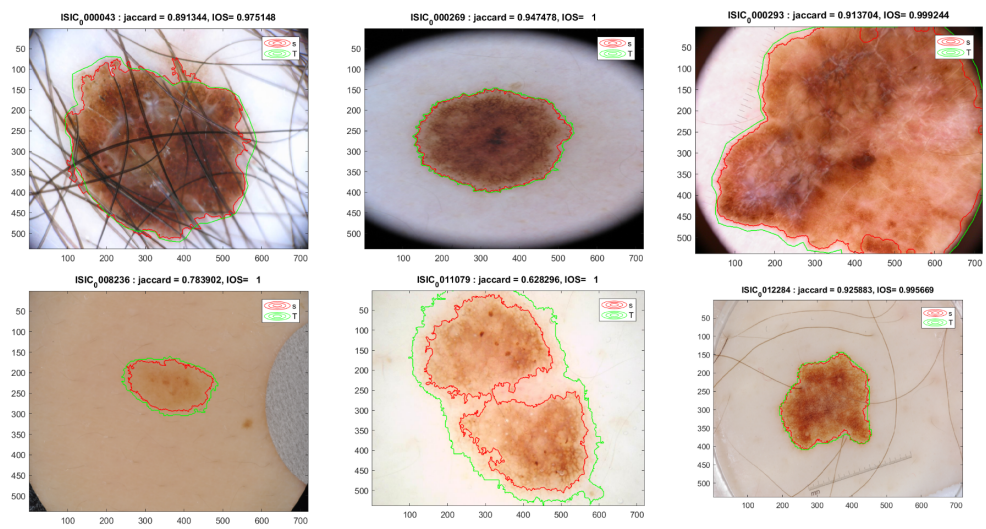


Figure 3: Good Results, T(green line) is label, S(red line) is segmentation

5.2.2 Bad Result

Figure 4 displays the bad results generated by the algorithm. Two evaluation criteria are both very low in these images, demonstrating the algorithm dis not perform a good segmentation of images

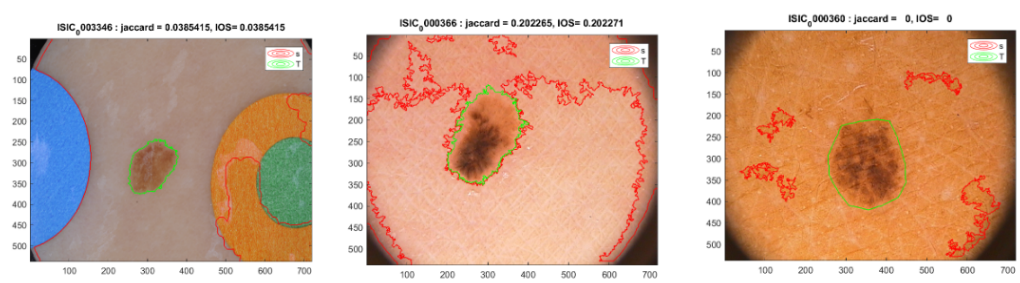


Figure 4: Bad Results, T(green line) is label, S(red line) is segmentation

5.3 Mask extracting results

The mask extracting is generated by thresholding the segmentation after local histogram equalization, a sample of generated mask in different level is shown in Figure 5

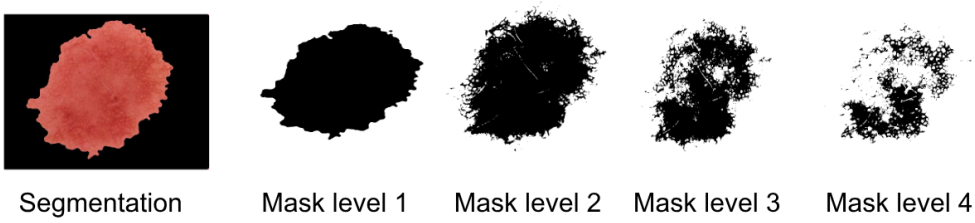


Figure 5: Masks in different levels of intensity value

5.4 Jaccard index

The overall performance of Jaccard index over 2594 segmentation labels are shown in the Figure 6

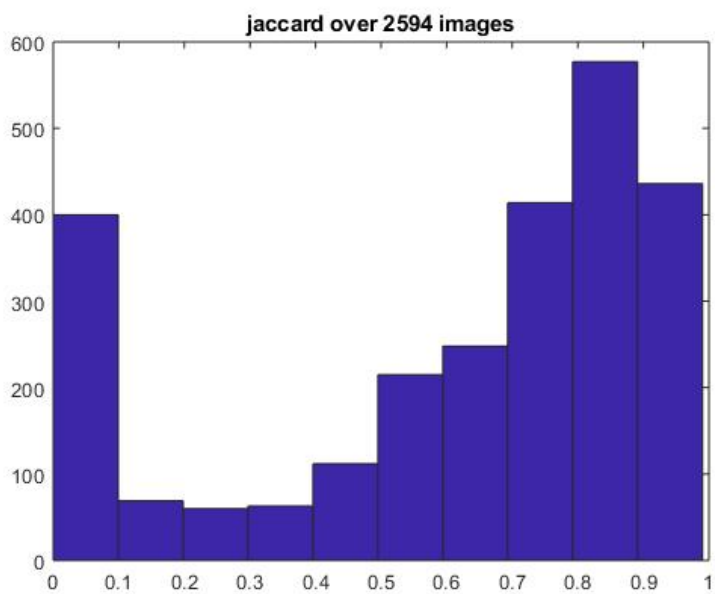


Figure 6: Jaccard index over 2594 segmentation

5.5 Ios index

The overall performance of Ios index over 2594 segmentation labels are shown in the Figure 7

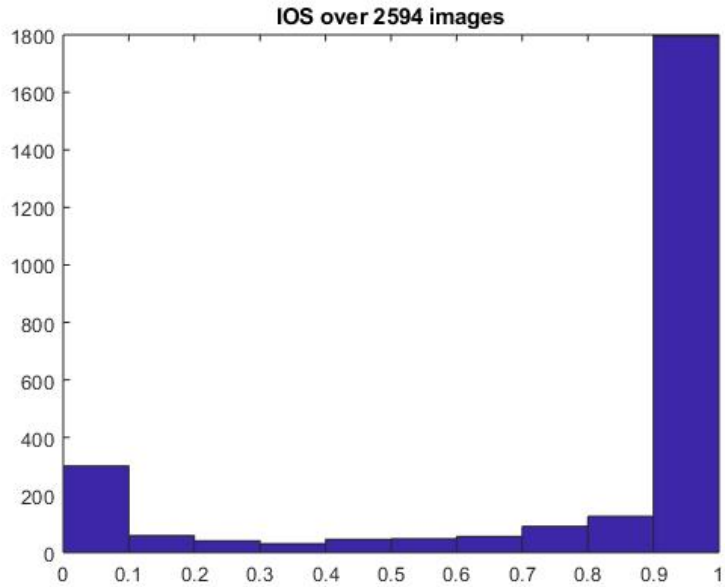


Figure 7: Ios index over 2594 segmentation

5.6 Dataset Analysis

The one major observation found about the ISIC challenge dataset during looking through the results, which is some segmentation labels are not accurate enough for shape analysis purpose. One Important reason Ios is introduce is some labels are found that not accurate enough, especially when the contrast between lesion area and skin area is not high enough and there are multiple skin areas can be viewed as cancer in the single image. The main reason behind this scenario is some labels are not generated by algorithm but human being. In another word, the human being can not segment the images very well so that some segmentation labels can not fully represent the shape of the skin cancer. Figure 8 demonstrates the cases that the segmentation results outperforms the segmentation labels. In these cases, the jaccard index always is not very high however the IOs is very high due to the fact that the segmentation are all within the area of the labels.

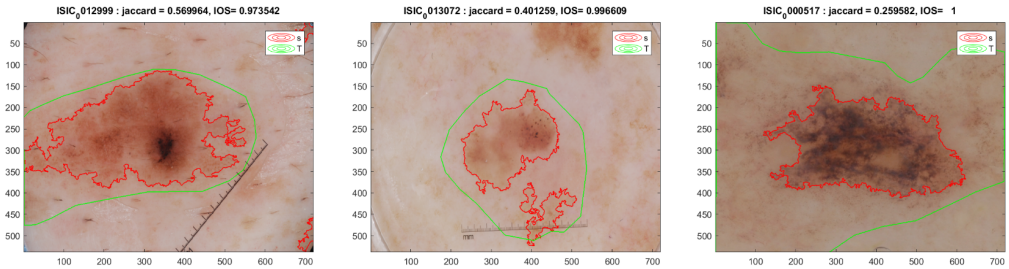


Figure 8: Cases that the segmentation results outperform the segmentation labels

Therefore when a segmentation is very unsatisfactory, both Jaccard and Ios indices are very low. Based on this conclusion, the two dimension histogram of Jaccard and Ios indices are used to better evaluate the performance of the algorithm, which is shown in Figure 9

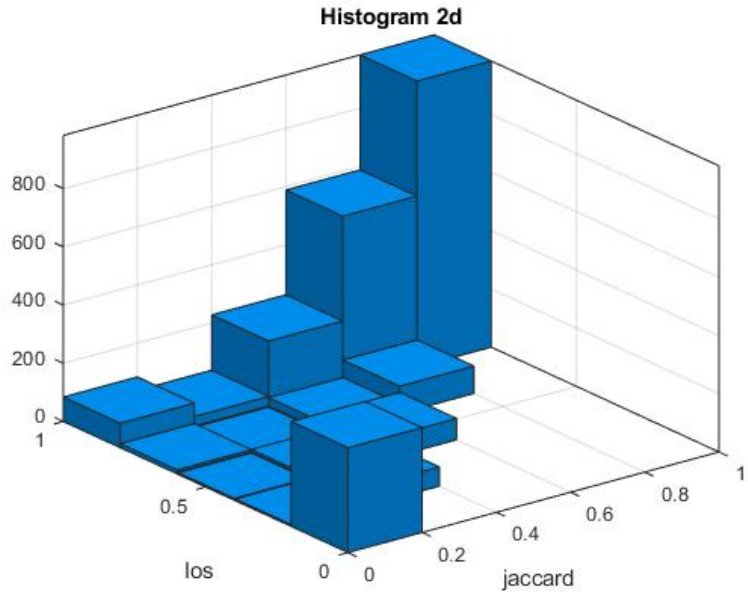


Figure 9: 2d histogram of Jaccard and Ios indices

It can be seen from the histogram that there are less than 250 segmentation results having very low Jaccard and Ios index (0-0.2), comparing with more than 900 segmentation results which have very high value (0.8-1.0) in both Jaccard and Ios index. This result can have a more comprehensive view of the algorithm performance.

5.7 Algorithm Analysis

There are some observations found about the algorithm during looking through the results. Some artifacts and unsatisfactory results generated by the algorithm imply the further optimization of the algorithm.

5.7.1 Bad performance when there are multiple dark areas in the image

When there are multiple dark areas in the input images, the algorithm sometimes gets confused and includes the non-cancer areas into the final segmentation results, as shown in Figure 10. The optimization could be done by having more conditions on selecting each connected components.

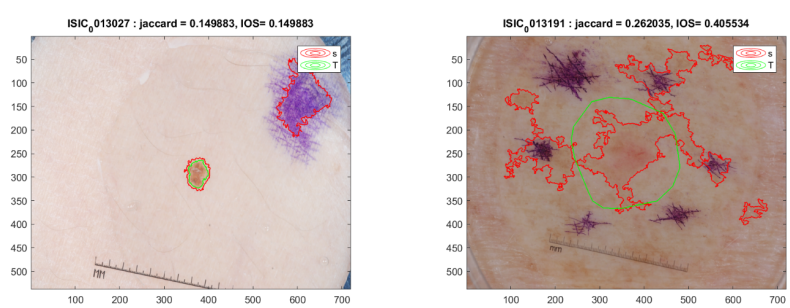


Figure 10: When there are multiple dark areas in the images the algorithm may not always select the correct cancer area

5.7.2 Bad performance on artifacts of preprocessing

The DullRazor hair removal algorithm and the scope shadow removal sometimes generate artifacts, which affect the segmentation results as shown in Figure 11. An improvement suggestion is to use a more aggressive dilation after the black shadow is detected and use more aggressive filter to filter out the hair line that is too board.

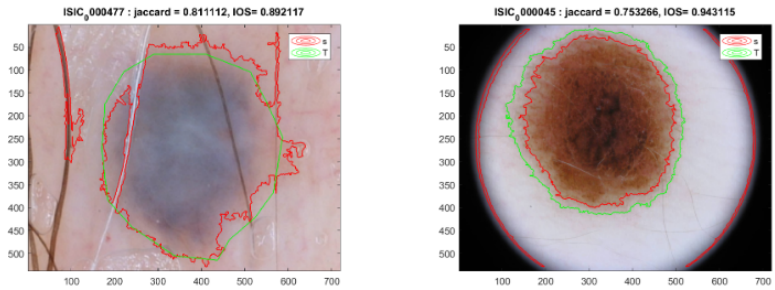


Figure 11: the artifacts generated by hair removal(left) and scope shadow removal(right)

6 Conclusion

In this project, a classic image processing algorithm is designed to segment the lesion picture for shape analysis. After demonstrating the dataset and method used in this project, performance analysis shows that the algorithm can have a even better performance on extracting shapes than the labels do. Finally, some improvement suggestions are proposed to have further optimization of the algorithm.

Acknowledgment

I would like to express my gratitude towards Professor Alon Orlitsky for his kindness help. I would also like to thank Professor Vwani Roychowdhury for his patient guidance.

References

[1] Tschandl, Philipp, 2018, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", https://doi.org/10.7910/DVN/DBW86T, Harvard Dataverse, V1, UNF:6:IQTf5Cb+3EzwZ95U5r0hnQ

[2] "Task 1: Lesion Boundary Segmentation," ISIC 2018. [Online]. Available: https://challenge2018.isic-archive.com/task1/. [Accessed: 22-Mar-2019].

[3] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. Mclean, "Dullrazor: A software approach to hair removal from images,"Computers in Biology and Medicine, vol. 27, no. 6, pp. 533â543, 1997.

[4] D. Liu and J. Yu, "Otsu Method and K-means,"2009 Ninth International Conference on Hybrid Intelligent Systems, 2009.