University of Girona

Spain

Medical Imaging Analysis
Lab 3 - Image Segmentation

Submitted by :
Mr. Shubham WAGH
Mr. Nayee Muddin Khan
DOUSAI

1 Introduction

Image segmentation is an important processing step in a Computer Aided Detection / Diagnsosis framework. It can used to focus further machine learning processing on a specific organ or lesion location. In this lab, we focus on dermatoscopic images in which the lesion in each image has to be delineated. For this lab, we deal with three typical problems found in segmentation of medical images:

1. Ground Truth generation.
2. Segmentation Evaluation
3. Segmentation Algorithm

We discuss each of these steps in detail and also present their implementation results.

2 Ground-truth generation

We are given 8 dermatoscopic images in which the lesion in each image has to be delineated. Three experts propose a ground-truth segmentation of the original images. Here, the first task is to fuse these different opinions to form a single ground-truth.

2.1 Implementation

Firstly, we visualized each of the segmented ground truths proposed by all the three experts and concatenated them in a single image for all the images. This helps us to visualize the overlapping areas from different experts for each segmented image.

Initially for the fusion of these segmented ground-truth, we take only the intersection areas

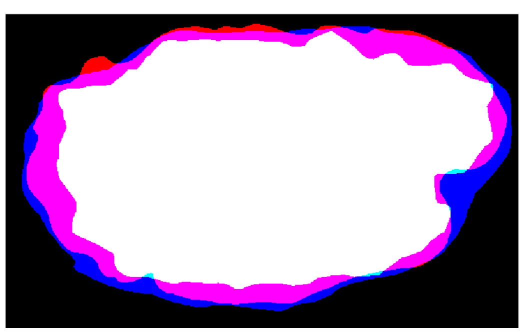

Figure 1: Concatenated image of *D134_mask* from the three experts. *White* area is the intersection areas of all the experts. **Pink** areas are the intersection of two experts. **Blue** and **Red** are the non overlapping areas from the experts

of the segmented ground-truths from all the three experts, as our first hand on fusion step. Although this step has its limitations, but it certainly provides us those areas of the lesion (single ground truth) for each image for which all the three experts agree or in other words, it gives us those areas of lesion where the segmented ground-truth proposed by the experts exactly matches with each other. As mentioned earlier, this intersection of ground-truths from all the three experts has its limitations :

1. Intersection of ground-truths can certainly give exact areas where lesion is, but this can skip those areas where there are in fact lesions. This is not acceptable.
2. Also intersection areas is only considered when all the three experts agree or the pixel value for all the 3 ground-truths for each image are equal. There can be some areas where two experts agree or the pixel value for two ground-truths are equal. This should be taken into consideration as well.

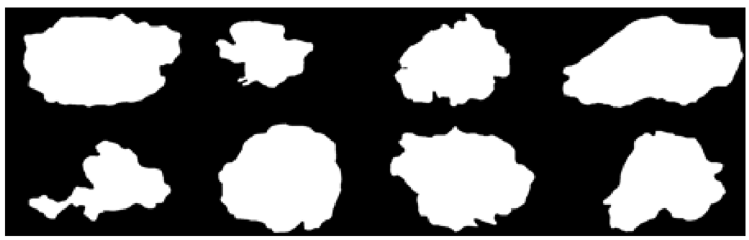

Figure 2: Ground-truth fusion taking the intersection areas

A new approach is required for fusion to a single ground-truth. We perform a majority voting based scheme or STAPLE which is given as reference in the lab guide.

2.1.1 STAPLE

STAPLE is an expectation-maximization algorithm for Simultaneous Truth And Performance Level Estimation. The algorithm considers a collection of segmentations and computes a probabilistic estimate of the hidden, implicit, true segmentation and a measure of the performance level achieved by each segmentation. The source of each segmentation in the collection may be an appropriately trained human rater or raters, or automated segmentation algorithms. The probabilistic estimate of the true segmentation is formed by estimating an

Figure 3: Ground-truth fusion using STAPLE

optimal combination of the segmentations, weighting each segmentation depending upon the estimated performance level, together with a prior model that can account for the spatial distribution of structures and spatial homogeneity constraints. It readily enables assessment of the performance of an automated image segmentation algorithm, and enables direct comparison of human rater and algorithm performance.

3 Segmentation

In this task, we are given three different segmentation algorithms to perform segmentation on 8 different dermoscopic images. We discuss the principle of these segmentation algorithms in brief.

3.1 PDF-based Segmentation

This method performs adaptive threshold based on local mimima. Firstly histogram of the input image is taken which gives us the counts of the intensity values. Then the histogram curves are smoothed which creates the data for a smooth line going through the points given. The maximum and minimum points are stored by specifying some *delta* value. A point is considered a maximum peak if it has the maximal value, and was preceded by a value lower by *delta*. Once all the minimums are sorted in ascending order, the first minimum value is taken as threshold otherwise if no mimimum is found, specified threshold value is used . This is used to create a mask for the input image which is used to find the number of largest connectivity and then the holes are filled just in case if any. This method uses probability density function, in our case histogram of the input image. Hence PDF-based segmentation which is based on local mimimum points (adaptive thresholding).

3.2 Level-set Segmentation

This method deals with intensity inhomogeneities in the segmentation. First, based on the model of images with intensity inhomogeneities, a local intensity clustering property of the image intensities is derived, a local clustering criterion function is defined for the image intensities in a neighborhood of each point. This local clustering criterion function is then integrated with respect to the neighborhood center which gives a global criterion of image segmentation. In a level set formulation, this criterion defines an energy in terms of the level set functions that represent a partition of the image domain and a bias field that accounts for the intensity inhomogeneity of the image. Therefore, by minimizing this energy, the method is able to simultaneously segment the image and estimate the bias field, and the estimated bias field can be used for intensity inhomogeneity correction (or bias correction).

$$F(\phi, c, b) = \int \sum_{i=1}^{N} \Big(\int K(y-x)|I(x)-b(y)c_i|^2 dy) M_i(\phi(x))dx + \mu \int p|\nabla \phi|dx + \nu \int |\nabla H(\phi)|dx \Big)$$

Parameters:
ν - coefficient of arc length term, μ - coefficient for distance regularization term (regularize

the level set function), σ - scale parameter that specifies the size of the neighborhood (used to control K), N - number of segmented materials.

3.3 Fuzzy C-means Segmentation

The FCM algorithm is one of the most widely used fuzzy clustering algorithms. The FCM algorithm attempts to partition a finite collection of elements $X =$, , ..., into a collection of c fuzzy clusters with respect to some given criterion. Given, a finite set of data, the algorithm returns a list of c cluster centers V, such that

$$V = v_i, i = 1, 2, \ldots\ldots, c$$

And a partition matrix U such that

$$U = u_{i,j}, i = 1, 2, \ldots\ldots, c, j = 1, 2, \ldots\ldots., n$$

where $u_{i,j}$ is a numerical value in $[0, 1]$ that tells the degree to which the element x_j belongs to the i^{th} cluster.

3.4 Otsu Segmentation

Here, we implement our own segmentation algorithm on 8 different dermoscopic images and find the ground truth for each of them, which is later used to evaluate its performance and precision. Firstly, we convert the dermoscopic image from RGB color space to HSV color space. The HSV color space abstracts color (hue) by separating it from saturation and pseudo-illumination. Also it is best suited for color based segmentation. Then we extract only the saturation component of all the images and apply MATLAB's inbuilt Otsu segmentation function to perform segmentation of those dermoscopic images. After this we apply morphological operations (closing and erosion) to get the final segmented result correctly (with less number of holes).

Figure 4: Segmentation using Otsu thresholding followed by morphological operation

4 Implementation and Evaluation Results

Once we generate all the segmented images using the above mentioned segmentation algorithms, we then evaluate them on the basis of segmentation metrics (Dice coefficient, Sensitivity, Specificity, Precision, Negative Productive Value and Hausdorff Distance). These segmentation metrics help us to evaluate the performance of segmentation algorithms. We discuss and report these results in this section.

4.1 Segmentation metrics

- **Sensitivity** - also named as True Positive Rate (TPR), measures the proportion of positives and it is described as the ratio of number of true positive to positive samples.
- **Specificity** - also named as True Negative Rate (SPC), measures the proportion of negatives and it is described as the ratio of number of true negative to negative samples.
- **Precision** - also named as Positive Predictive Rate (PPV) and described as ratio of true positive to sum of true and false positives.
- **Negative Predictive Value** - it is described as the ratio of true negative to the sum of true and false negatives.
- **Dice coefficient:** - it is a measure for comparing the similarity of two images.
- **Hausdorff Distance:** - it measures the degree of mismatch between two sets by calculating the distance of point in one set that is the farthest from any point in other set, or vice-versa.

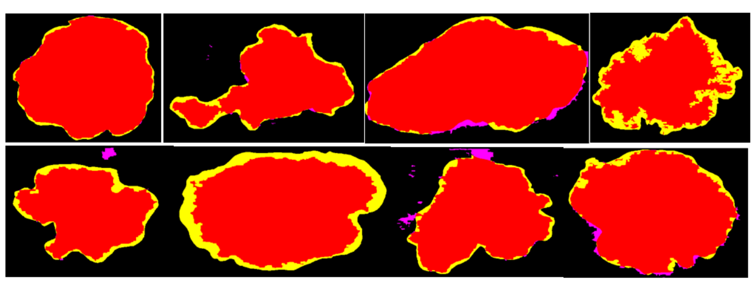

Figure 5: Otsu Segmentation result against fusion ground-truth images. **Red** represents segmentation result. **Yellow** represents ground-truth. **Pink** represents areas which are not part of lesion

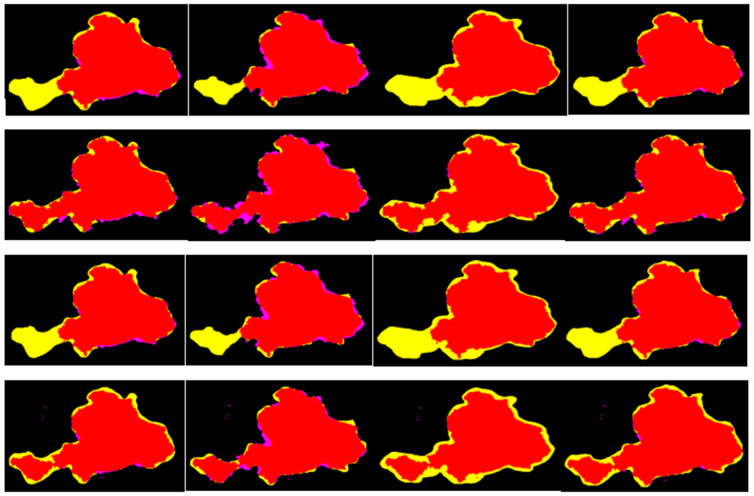

Figure 6: Segmentation comparison results for E976_mask Image. **Top to Bottom**: FCM, Level-Set, PDF and Otsu Segmentation. **Left to Right**: Expert 1, Expert 2, Expert 3 and Fusion ground-truth Images

4.2 Segmentation Algorithm Evaluation

In this section we evaluate the performance of the above mentioned segmentation algorithms including Otsu segmentation (our implementation) based on the segmentation metrics discussed in the previous section. We evaluate the algorithm performances on the basis of two categories:

- **Region based measures** - Dice coefficient, Sensitivity, Specificity, Precision and Negative Predictive value.
- **Boundary based measures** - Hausdorff distance.

4.2.1 Region based measures

Here we evaluate the images generated by the above mentioned segmentation algorithms against the fusion ground-truth images using STAPLE on the basis of region based measure parameters. And this is done for all 8 images. We then take average of each parameter for each image. Then add all the parameters together to get a final region based final score. We then take the maximum score from all the segmentation algorithms. The one with the maximum score is the better segmentation algorithm.

4.2.2 Boundary based measure

Here we take the average of Hausdorff distance for each image for different segmentation algorithms. The one with the minimum Hausdorff distance is the better segmentation algorithm. Here we take the minimum score as Hausdorff distance is a measure of mismatch between two images.

	FCM		Level-Set		PDF		OTSU	
Fusion Images	Dice coefficient	0.81308	Dice coefficient	0.93247	Dice coefficient	0.91392	Dice coefficient	0.93229
	Specificity	0.98200	Specificity	0.9937	Specificity	0.98291	Specificity	0.86722
	Precision	0.98232	Precision	0.98976	Precision	0.98675	Precision	0.98766
	Negative Predictive Value	0.91014	Negative Predictive Value	0.89862	Negative Predictive Value	0.91459	Negative Predictive Value	0.91068
	Sensitivity	0.89017	Sensitivity	0.88530	Sensitivity	0.85862	Sensitivity	0.87826
	Hausdorff distance	90.2776	Hausdorff distance	40.8895	Hausdorff distance	54.9484	Hausdorff distance	67.9989

	FCM	Level-Set	PDF	OTSU
Region Based Measure Score	0.91554	0.93997	0.93135	0.91522
Boundary Based Measure Score	90.2776	40.8895	54.9484	67.9989

Figure 7: Segmentation Evaluation against fusion ground-truth images. Table shows the average scores for each parameter for all the images for different segmentation algorithms

From the above table we can conclude that *Level-set segmentation* gives us the maximum region based score (0.93997) also minimum boundary based score (40.8895). This shows that *Level-set segmentation* algorithm gives us better results as compared to other segmentation algorithms. Otsu-segmentation ranks third as per our evaluation criteria.

5 Conclusion

We conclude that *Level-set* segmentation algorithm performs better as compared to other segmentation algorithms and gives better results as per our evaluation criteria. We also learnt about different segmentation metrics namely region based measures and boundary based measures. This lab solidified our knowledge regarding segmentation of medical images, different segmentation algorithms and segmentation metrics for performance evaluation.