Automated Skin Lesion Segmentation via Image-wise Supervised Learning and Multi-Scale Superpixel Based Cellular Automata

Lei Bi¹, Jinman Kim¹, Euijoon Ahn¹, Dagan Feng¹,4 and Michael Fulham¹,2,3

¹School of Information Technologies, University of Sydney, Australia
²Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Australia
³Sydney Medical School, University of Sydney, Australia
⁴Med-X Research Institute, Shanghai Jiao Tong University, China

ABSTRACT
Segmentation of skin lesions is considered as an important step in computer aided diagnosis (CAD) for automated melanoma diagnosis. Existing methods however have problems with over- or under-segmentation and do not perform well when a lesion is partially connected to the background or when the image contrast is low. To overcome these limitations, we propose a new automated skin lesion segmentation method via image-wise supervised learning (ISL) and multi-scale superpixel based cellular automata (MSCA). We propose using ISL to derive a probabilistic map for automated seeds selection, which removes the reliance on user-defined seeds as in conventional methods. The probabilistic map is then further used with the MSCA model for skin lesion segmentation. This map enables the inclusion of additional structural information and when compared to single-scale pixel-based CA model, it produces higher capacity to segment skin lesions with various sizes and contrast. We evaluated our method on two public skin lesion datasets and showed that it was more accurate and robust when compared to the state-of-the-art skin lesion segmentation methods.

Index Terms—Segmentation, Cellular Automata, Skin lesion, Melanoma

1. INTRODUCTION
Malignant melanoma has one of the most rapidly increasing incidences in the world and it causes considerable mortality [1]. Early diagnosis is particularly important since melanoma can be cured with excision if detected early [2]. Even for experienced dermatologists, however, diagnosis by human vision can be subjective, inaccurate and non-reproducible [1]. This is attributed to the complexity of lesion segmentation due to variations in size and shape, fuzzy lesion boundaries and different skin colors. Motivated by these difficulties, there has been a great interest in developing computer-aided diagnosis (CAD) systems that can assist the dermatologists’ clinical evaluation [1, 3, 4].

Lesion segmentation is a fundamental requirement for a melanoma CAD. A number of segmentation methods have been proposed recently to automatically segment skin lesions. These methods include saliency based skin lesion segmentation (SSLS) [5], deformable model (DM) [1], texture distinctive lesion segmentation (TDLS) [3], adaptive thresholding (AT) [4], level set proposed by Chan et al (CLS) [6] and seeded region growing [7]. However, these methods usually rely on manual initialization [1], assume lesions are completely inside the images [3, 5] and the images are illumination corrected [3]. In addition, these methods cannot provide accurate segmentation when lesions have complex shapes and are low-contrast to the background [4, 6].

A cellular automata (CA) model has been proposed for many segmentation problems in medical images [8-10]. The CA model is based on a lattice of ‘cell’ (pixels), which are assigned to the foreground (lesion), background or undefined. For each iteration of the algorithm, cells propagate across the whole lattice according to the cell’s features. CA is tolerable to image pattern complexity and low-contrast issues [8-10]. However, existing applications of CA rely on using pixel-level information and often require operator-dependent manual or supervised learning approaches to initialize the seeds selection. Qin et al [11] recently reported the combined use of CA with superpixels for saliency detection on general images. This combination has high accuracy for detecting the foreground (saliency area) due to its ability to derive additional structural information from superpixels. However, its reliance on a single scale superpixel may limit its ability to detect foreground objects that have varying sizes. In addition, its application of a linear parametric model for seed selection means that it is dependent on a priori knowledge and thus is not able to be optimized for individual input images. For such a CA and superpixel combination to work on skin images, it must be adapted to work on different skin lesion sizes; be able to derive image specific parametric models to cater for large visual variations among different studies; and to be optimized for the skin imaging characteristics.

In this study, we propose a new automated skin lesion segmentation method via image-wise supervised learning (ISL) and multi-scale superpixel based cellular automata (MSCA). The novelty of our algorithm when compared to previous studies is as follows: (1) we propose an image-wise supervised learning approach to initialize seeds via a
probabilistic map between the skin lesion and the background. This initialization improves the capacity for the following CA to segment the lesion area, whereas traditional methods usually require CA to propagate from user defined or predefined seeds; and (2) we propose a novel multi-scale superpixel based CA model with a parallel propagation.

2. METHODS

2.1. Pre-processing

Artifacts such as hair can degrade the performance of the segmentation methods. The presence of hairs on the skin may occlude parts of the lesion, making accurate skin lesion segmentation difficult. Hence, we adopted the hair removal algorithm reported by Lee et al [12] to remove visible hair in the images. Hair was detected by using multiple hair templates; the detected hair pixels were replaced with non-hair pixels derived from neighborhood pixels. We also detected dark background by finding dark superpixels connected to the boundaries (see Section 2.2). These detected superpixel regions were also replaced with neighborhood non-boundary pixels.

2.2. Image-wise Supervised Learning based Initialization

![Figure 1. Segmentation results at different stages. (a) input image; (b) superpixel results; (c) image boundary clustering results, where different colors represent different clusters; (d-f) probabilistic maps based on different boundary clusters; (g) averaged probabilistic map of (d-f); and (h) cellular automata processed results (Section 2.3).]

An input image is segmented into N small superpixels by the linear spectral clustering [13] algorithm (Fig. 1b). To achieve an optimal probabilistic map for seed selection (lesion/foreground area), we applied the K-means algorithm to partition the image boundary (boundary superpixels) into k clusters based on their mean CIE LAB color features (Fig. 1c). We empirically set the number of boundary clusters to $k = 3$. The boundary superpixels belonging to cluster $k$ is then represented as $C^k = \{s^k_1, s^k_2, ..., s^k_p \}$, where $s^k_p$ represents the superpixel and $p^k$ represents the number of superpixels in cluster $k$. After that, for each cluster $k$, we constructed the initial probability map which represents the possibility of a pixel falling into the lesion area. This was achieved by measuring the similarity of all non-boundary superpixels to the boundary clusters. Initially, we divided all the boundary clusters into two sets $D_1$ and $D_2$, where $D_1$ was set equal to $C^k$ and $D_2$ represents the rest of boundary superpixels. We labelled all the superpixels in $D_1$ as $+1$ and $D_2$ as $-1$. A binary classifier $C$ was then learned to separate the two sets ($D_1$ and $D_2$). After that, we tested the classifier $C$ on the rest of the superpixels in the image. The probability score (derived from classification) of each superpixel represents the similarity of that superpixel to the background. We iteratively generated $k$ different probabilistic maps by using different clusters (Fig. 1d-f). In general, the $k$ different probabilistic maps complement each other and such an example is given in Fig. 1. In this example, in Fig. 1d the map (derived from the light blue cluster in Fig. 1c) can separate the background while Fig. 1e (derived from the red cluster) only separated the lesion boundary. We summed all different maps to minimize the risk of selecting the wrong map (Fig. 1g). In this study, the binary classifier $C$ was set to be a support vector machine (SVM) with a radial basis function (RBF) kernel. A RBF kernel was used to non-linearly map the data into a higher dimensional space [14]. This helps make the training data more identifiable in a non-linear classification task [15]. The RBF kernel parameters were optimized with a default grid search and analysis method, which is available in the LIBSVM [14].

2.3. Cellular Automata Based Segmentation

The cellular automata is a dynamic system. The system consists of a lattice of cell with discrete states, which evolve in discrete time steps according to definite rules [9]. Each cell’s next state will be determined by its current state and the states of its nearest neighbors. However, discrete states cannot sufficiently represent the probability of the lesion pixels since it can only represent two states (0 or 1). We followed the modification used by Qin et al [11] to have continuous states, which can be defined as:

$$S^{t+1} = C \cdot S^t + (1 - C) \cdot F \cdot S^t$$

(1)

where $F$ is a normalized weight matrix ($N \times N$) of $F^t$. $F^t$ defines the similarity of two superpixel using Euclidean distance and we normalized $F^t$ according to the sum of each row. We used the same superpixel color features mentioned in Section 2.2 to measure the similarity. I is the identify matrix with ones on the diagonal and zeros elsewhere. $C$ is a normalized coherence matrix of $C^t$ with normalized value on the diagonal and zeros elsewhere. $C^t$ is defined as $C^t = diag(c_{1}, c_{2}, ..., c_{N})$ and $c_{j}(1 \leq i \leq N)$ is calculated as:

$$c_{i} = a \cdot \frac{c_{j} - \min(c_{j})}{\max(c_{j}) - \min(c_{j})} + b$$

(2)

where $j$ is the neighborhood superpixel of $i$ and $c_{j}$ is defined as multiplicative inverse of the maximum difference between the neighborhood superpixels to $i$. $a, b$ are two constant values were empirically set to 0.6 and 0.2.

$S^t$ from equation (1) defines the current probabilistic map after the $t$ number of iterations. For $t = 0$, we used the probabilistic map generated in Section 2.2. By analyzing the superpixel with their surrounding neighborhood and trying to push all the superpixels into a stable state, CA could effectively increase the probabilistic of the foreground while
reducing the probabilistic elsewhere. Fig. 1 shows an example where the initialization map is not feasible (1g). The result, however, after CA is still satisfying (1h). In this study, the number of iterations was empirically set to 20.

2.4. Multi-scale Integration and Post-refinement

Figure 2. CA generated probabilistic maps on each scale, where (a) is the input image, (b-g) are the generated probabilistic maps using 50 – 300 number of superpixels and (h) final integrated segmentation result.

To calculate the final lesion segmentation, we first integrated the probabilistic maps across all different scales to achieve pixel level probabilities, through a weighted average which can be defined as:

$$
\phi(q) = \frac{\sum_\sigma P^\sigma(q)}{\sum_\sigma \omega^\sigma} 
$$

(3)

where $q$ is a pixel, $P^\sigma$ represents the probability at scale $\sigma$ and $\omega$ is the associated weight. It is expected that we are more confident about our probabilistic map if it has a good separation to the background. In this case, we should approach a higher weight to that probabilistic map. To measure and compare this separation, we conducted a histogram analysis on each scale to count the number of empty bins and we used this number to derive the weight, which can be defined as:

$$
\omega^\sigma = e^{\left| \mu(\sigma) - \min_{\delta \in Y} \mu(\delta) \right|} 
$$

(4)

where $\mu$ counts number of empty bins and $Y$ represents the number of different scales. To cover lesions of various sizes, we used 10 different scales ranging from superpixel number (in an image) of 50 to 500 at increments of 50.

The final integrated probabilistic map was converted into a binary segmentation result via a thresholding method proposed by Li et al [16]. For segmentation refinement, we followed previous work of Glaister et al [3] and used a morphological dilation process with a disc radius of 5 pixels to smooth the boundary, fill small holes and use connected thresholding to remove small isolated single pixels.

3. RESULTS AND DISCUSSIONS

3.1. Materials and Experimental Setup

Two public image datasets were used to test the effectiveness of our algorithm: (i) based on [1], 160 dermoscopy images were selected from the PH2 dataset [17]; (ii) 106 digital photographs images were selected from the Dermquest dataset [18] and the following images were excluded: (a) images having multiple lesions with the ground truth only annotated one; (b) images that were heavily affected by illumination; and (c) hair or the background occupying more than a 1/3 of an image. Manually annotated lesions from dermatologists (available for both datasets) were used as the ground truth.


We used the preferred evaluation metrics published in the original literature [1, 3, 5] for benchmarking including: (i) dice similarity coefficient (DSC) [5] – measures the overlap between ground truth (denoted as $g$) and the algorithm produced results (denoted as $a$); (ii) $D_0$ [1] – measures the percentage of segmentation errors; (iii) $D_1$, $D_2$, and $D_1 + D_2$ [1] – measures the average minimum point-to-contour distance ($g$ to $a$) and ($a$ to $g$) and their sum; (vi) sensitivity (Sens.), specificity (Spec.) and accuracy (Accu.) [3]; and (ix) balanced accuracy (BAC) of the segmentation results which is defined as (Sens+Spec)/2 to balance the importance of sensitivity and specificity.

3.2. Results

Table 1 shows that our method was the best performed across all DSC measurements on the PH2 dataset and is also had 3 of the 4 lowest point-to-contour measurements. Table 2 shows that our method had the best accuracy and the best balanced accuracy (BAC); it was 2nd best for sensitivity and specificity (0.52% lower to the best).

Table 1: Segmentation results on the PH2 dataset with our method compared to the other comparison methods. Red and blue represent the best and the second best performing methods, respectively.

<table>
<thead>
<tr>
<th>Method</th>
<th>AT</th>
<th>C-LS</th>
<th>SRG</th>
<th>SSLS</th>
<th>DM</th>
<th>Our</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSC-Avg (%)</td>
<td>82.36</td>
<td>87.02</td>
<td>61.24</td>
<td>91.33</td>
<td>-</td>
<td>92.49</td>
</tr>
<tr>
<td>DSC-Min (%)</td>
<td>11.14</td>
<td>6.99</td>
<td>0.22</td>
<td>40.30</td>
<td>-</td>
<td>75.19</td>
</tr>
<tr>
<td>DSC-Max (%)</td>
<td>97.36</td>
<td>97.87</td>
<td>97.20</td>
<td>97.38</td>
<td>-</td>
<td>98.11</td>
</tr>
<tr>
<td>DSC-Std (%)</td>
<td>14.96</td>
<td>11.24</td>
<td>30.43</td>
<td>7.27</td>
<td>-</td>
<td>4.09</td>
</tr>
<tr>
<td>D0-Avg (%)</td>
<td>41.09</td>
<td>26.76</td>
<td>50.23</td>
<td>17.48</td>
<td>13.92</td>
<td>14.99</td>
</tr>
<tr>
<td>D1-Avg</td>
<td>27.79</td>
<td>19.35</td>
<td>43.22</td>
<td>12.33</td>
<td>10.41</td>
<td>10.40</td>
</tr>
<tr>
<td>D2-Avg</td>
<td>19.35</td>
<td>12.91</td>
<td>49.29</td>
<td>12.28</td>
<td>9.77</td>
<td>7.65</td>
</tr>
<tr>
<td>D1-Avg+D2-Avg</td>
<td>47.14</td>
<td>32.26</td>
<td>92.51</td>
<td>24.61</td>
<td>20.18</td>
<td>18.05</td>
</tr>
</tbody>
</table>

Table 2: Segmentation results of our method when compared to others on the Dermquest dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>L-SRM</th>
<th>Otsu-R</th>
<th>Otsu-RGB</th>
<th>Otsu-PCA</th>
<th>TDLS</th>
<th>Our</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sens (%)</td>
<td>89.40</td>
<td>87.30</td>
<td>93.60</td>
<td>79.60</td>
<td>91.20</td>
<td>93.08</td>
</tr>
<tr>
<td>Spec (%)</td>
<td>92.70</td>
<td>85.40</td>
<td>80.30</td>
<td>99.60</td>
<td>99.00</td>
<td>99.08</td>
</tr>
<tr>
<td>Acce (%)</td>
<td>92.30</td>
<td>84.90</td>
<td>80.20</td>
<td>98.10</td>
<td>98.30</td>
<td>98.80</td>
</tr>
<tr>
<td>BAC (%)</td>
<td>91.05</td>
<td>86.35</td>
<td>86.95</td>
<td>89.60</td>
<td>95.10</td>
<td>96.08</td>
</tr>
</tbody>
</table>

3.3. Discussion

Table 1 results confirm that our method performed the best in segmenting the lesions in the PH2 dataset. Although SSLS achieved competitive results in the DSC measures, it failed
to separate the lesion when it was partially connected to the boundary. AT, C-LS and SRG consistently suffered from low-contrast and inhomogeneous areas (Fig. 3). The DM result was the 2nd best when compared to our method on point-to-contour measures. However, it required user input to initialize the model and this manual input is operator-dependent and image specific which prevents it being considered in an automated CAD system.

Table 2 shows that our method performed better compared to the other methods on the Dermquest dataset. This dataset benchmarks our method to new skin lesion image characteristics. When compared to dermoscopy images, digital photographs introduce more challenges as they usually have low contrast and non-uniform illumination conditions [3]. The results show that our method had the best results. Compared with TDLS (the 2nd best performing method), our method had higher sensitivity (~2%), which implies better detection of true positive pixels (lesions). This was likely due to the multi-scale processing; where multi-scale processing could identify scale invariant structure which usually the lesions and approach a higher probabilistic. TDLS relied on a supervised learning approach (a priori knowledge) to initialize the segmentation and it only worked with illumination corrected images. In contrast, our method does not require priori knowledge and is insensitive to the illumination conditions.

![Figure 3. Segmentation results from four example studies.](image)

The segmentation performance using CA with a single-scale superpixel is shown in Fig. 4. It shows that the inclusion of the multi-scale superpixel levels give superior performance when compared to any single-scale levels. It also indicates that a larger superpixel scale (e.g., 50 superpixels in an image) has higher segmentation accuracy compared to a smaller superpixel scale (e.g., 500). This implies that pixel-level CA has lower segmentation accuracy, which can be attributed to the larger superpixels (e.g., 50) constituting more discriminative features that are crucial for lesion detection. Meanwhile smaller superpixel scales can only represent a small part of the lesion area. Nevertheless, our experiments show that small superpixel scales are still useful for multi-scale integration and could contribute to the precise boundary definitions.

![Figure 4. Segmentation performance with single-scale superpixels (blue) (e.g., 500 superpixels in an image) and multi-scale superpixels (orange) evaluated on the PH2 dataset.](image)

4. CONCLUSION AND FUTURE WORK

We propose a new automated skin lesion segmentation method with image-wise supervised learning and multi-scale superpixel based cellular automata. Our experiments on two public skin lesion datasets showed that our method achieved better segmentation performance when compared to current state-of-the-art methods.

REFERENCES