Pattern Recognition 46 (2013) 1012–1019

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm

Fengying Xie [a,*], Alan C. Bovik [b]

[a] School of Aeronautics and Astronautics, Beihang University, Beijing 100191, China
[b] Department of Electrical and Computer Engineering, The University of Texas at Austin, TX 78712, USA

ARTICLE INFO

Article history:
Received 23 August 2011
Received in revised form
17 August 2012
Accepted 19 August 2012
Available online 30 August 2012

Keywords:
Dermoscopy images
Self-generating neural network
Image clustering
Automatic segmentation
Generic algorithms

ABSTRACT

A novel dermoscopy image segmentation algorithm is proposed using a combination of a self-generating neural network (SGNN) and the genetic algorithm (GA). Optimal samples are selected as seeds using GA; taking these seeds as initial neuron trees, a self-generating neural forest (SGNF) is generated by training the rest of the samples using SGNN. Next the number of clusters is determined by optimizing the SD index of cluster validity, and clustering is completed by treating each neuron tree as a cluster. Since SGNN often delivers inconsistent cluster partitions owing to sensitivity relative to the input order of the training samples, GA is combined with SGNN to optimize and stabilize the clustering result. In the post-processing phase, the clusters are merged into lesion and background skin, yielding the segmented dermoscopy image. A series of experiments on the proposed model and the other automatic segmentation methods (including Otsu's thresholding method, *k*-means, fuzzy *c*-means (FCM) and statistical region merging (SRM)) reveals that the optimized model delivers better accuracy and segmentation results.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Malignant melanoma (MM), the most deadly form of skin cancer, is one of the most rapidly proliferating cancers in the world, with an estimated annual incidence of 70,230 and 8790 deaths in the United States in 2011 [1]. In China, the incidence of MM has increased 3%–8% annually and has doubled over the past decade [2]. The earlier the diagnosis, the lower the metastatic risk: investigations have shown that the cure rate is nearly 100% if the skin cancer is recognized early enough and treated surgically [3].

Advances in dermoscopy (skin-surface microscopy or dermatoscopy) technology have contributed significantly to improved detection and survival rates [4,5]. Dermoscopy [6] is a non-invasive technique that combines optical magnification and liquid immersion with angle-of-incidence lighting or crosspolarized lighting to make the contact area translucent, consequently revealing subsurface structures of the skin. Dermoscopy yields 10%–27% higher sensitivity than clinical diagnosis, significantly improving the accuracy of dermatologists when diagnosing melanoma [4,5]. Yet, dermoscopic diagnosis remains subjective and is therefore associated with poor reproducibility. Because of this there has been a significant increase in interest in the development of automatic digital dermatoscopic image analysis methods over the last decade. Such processes typically consist of four stages: image acquisition, lesion segmentation or border detection, feature extraction, and classification. The segmentation stage is quite important, since it affects the accuracy of the subsequent steps. However, segmentation is quite difficult because [7]: (*i*) the transition between the lesion and the surrounding skin is usually of low contrast; (*ii*) the lesion borders are usually irregular and fuzzy; (*iii*) complicating artifacts are often present such as skin texture, air bubbles and hairs; and (*iv*) the interior of the lesion may exhibit variegated coloring.

To address these problems, a number of dermoscopic segmentation algorithms have been developed [8]. For convenience, we broadly classify these into three categories: thresholding, edge/contour-based and region-based. An effective thresholding method proposed by Grana et al. [9] uses Otsu's threshold to automatically segment the melanoma image, then selects *k* points for spline-based interpolation, yielding a smoothed lesion border. Thresholding methods such as this can achieve good results when there is good contrast between lesion and skin, but encounter problems when the modes of the two regions overlap. Edge/contour-based approaches were used in [10,11]. Rubegni et al. [10] segmented dermoscopy images using the zero-crossings of a LoG edge operator, while Zhou et al. [11] used an improved snake model to detect lesion borders. Edge and contour-based approaches perform poorly when the boundaries are not well defined, for instance when the transition between skin and lesion

* Correspondence to: Image Processing Center, Beihang University, Xueyuan road No.37, Haidian District, Beijing, 100191, China. Tel.: +86 15810672928; fax: +86 10 82338798.
E-mail address: cherryfyxie@gmail.com (F. Xie).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2012.08.012

is smooth. In such situations, the edges have gaps and the contour may leak through them. Region-based approaches have also been used. Some examples include multi-scale region growing [12], fuzzy *c*-means based on anisotropic mean shift [13], multi-resolution markov random fields [14] and statistical region merging (SRM) [7]. Region-based approaches have difficulties when the lesion or the skin region are textured, or have different colors present, which can lead to over-segmentation.

With the increasing availability of methods for segmenting dermoscopic images of skin lesions, the relative performances of the various models are of interest. In [15], four widely used color clustering algorithms were compared: median cut, *k*-means, fuzzy *c*-means and mean shift, without employing any spatial constraint. The mean shift algorithm gave the best results. In [16], six methods were compared: gradient vector flow (GVF), a level set method of Chan (C-LS), adaptive thresholding (AT), adaptive snakes (AS), EM level sets (EM-LS), and a fuzzy-based split-and-merge algorithm (FBSM). The authors concluded that the best semi-supervised methods are AS and EM-LS, while the best fully automatic method is FBSM.

Color is a significant feature for image segmentation and unsupervised color clustering has been successfully used for region-based segmentation [17]. Such data-driven methods have great potential for dealing with varied imaging situations, provided that an accurate model that is flexible enough to span the space of possible lesion image environments can be found. Since modeling such a high-dimensional complex space of possibilities is quite difficult, learning-based methods that can be trained on large datasets are of interest. Towards this end, we study and develop a color clustering model for dermoscopic images that combines the technique of the self-generating neural network (SGNN) [18] with genetic algorithms (GA). Using a measure of cluster validity, the clustering algorithm that we develop automatically determines an appropriate number of clusters. By merging the clustering regions into lesion and background skin, segmentation of dermoscopic images is achieved. When compared with other segmentation algorithms that use Otsu's thresholding method, *k*-means, fuzzy *c*-means and SRM, our model is shown to deliver high-quality segmentation results.

2. Self-generating neural networks (SGNN)

We briefly describe the learning tool that we will use. SGNN was developed in 1992 [18] using the idea of self-organizing maps (SOM) implemented within a self-generating neural tree (SGNT) architecture. It was studied in depth in [19,20], and is characterized by simplicity in network design, and speed of learning and self-organizing capability. As such, it is a good choice to learn clustering or classification with high performance [21].

As shown in Fig. 1, the SGNN can be implemented as a self-generating hierarchical neural tree (SGNT). Fig. 1(a) depicts a clustering sample set, where $e_i, i = A,B,\cdots E$ are the sample attributes. Fig. 1(b) is the generated SGNT following the SGNT

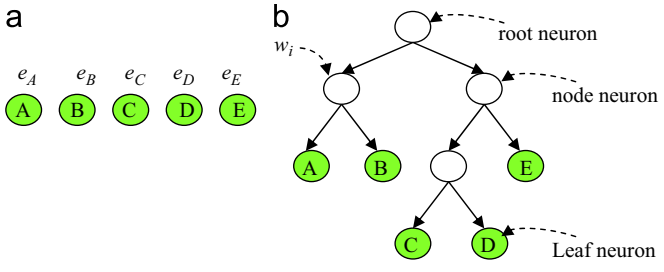

Fig. 1. Structure of the SGNT. (a) 5 samples and (b) Generated SGNT for (a).

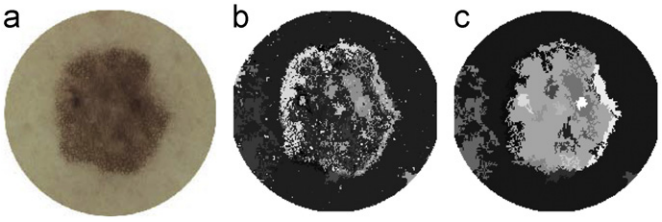

Fig. 2. Coarse segmentation with region growing. (a) Original image, (b) Region growing and (c) Filtering small sub-regions.

generating rules [18–21], where w_i notes neuron weight. Each leaf neuron corresponds to one or multiple samples, and its weight is the average attribute of the corresponding samples. The weight of every node neuron (non-leaf neuron) is the average attribute of all the leaf neurons it covers. Taking each child of the root neuron as a cluster center, each leaf neuron in the sub-network rooted by this child belongs to the same cluster. The number of clusters is consequently equal to the number of the root neuron's children. In Fig. 1(b), *A* is in the same cluster as *B*, and *C* is in the same cluster as *D* and *E*, while the number of clusters in the SGNT is 2. By taking image pixel values as the data, and color or location information as sample attributes, the SGNN can be used for image clustering.

For image segmentation purposes, the SGNT structure becomes excessively large if all pixels are trained. To reduce the complexity, we deploy a coarse-to-fine segmentation strategy. The region growing method is used to coarsely segment the original image. The image is scanned and the unlabeled pixels are taken as seeds. The pixel neighbors whose mean color has a distance less than 15 to the seed pixel color, are added to the region. Fig. 2(b) is the sub-regions segmented by region growing method, and very small regions are removed in Fig. 2(c); the size of these can be taken to be a small faction of typical lesion size. Using these sub-regions as the data set, the clustering task is well adapted to learning by the SGNN.

3. Automatic segmentation based on SGNN and GA

In spite of its fitting capacity for clustering, the SGNN algorithm is influenced by the input order of the training samples, which can cause inconsistent clustering results, as depicted in Fig. 3, where Fig. 3(a) is the original image, and Fig. 3(b) and (c) are the results that are arrived at when different samples are selected as the first input into the SGNT.

In Fig. 3(b), the area of the lesion is under-segmented, whereas a more accurate result is obtained in Fig. 3(c). To ameliorate this, we propose an adaptive clustering algorithm, termed ACluster-GA-SGNN, wherein the SGNT is generalized to a Self-Generating Neural Forest (SGNF), and GA is subsequently employed to consistently select an appropriate group of seed samples as the first input into the SGNF, thereby yielding optimized clustering results.

3.1. Self-generating neural forest

The SGNT can be generalized to a Self-Generating Neural Forest (SGNF) as follows. Suppose that a given sample set has *c* cluster centers. Then the SGNF generating algorithm can be described as follows:

Step 1: Remove *c* seed samples randomly from the sample set, treating these seeds as initial neural trees to form an initial forest.

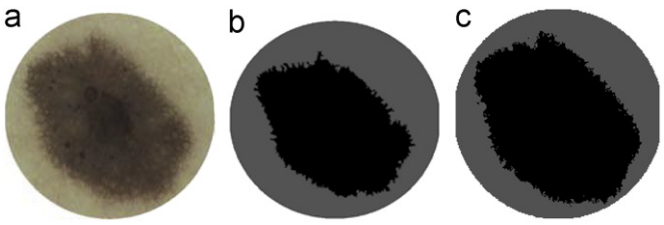

Fig. 3. Different clustering results using SGNN with different input order of samples. (a) Original image, (b) Clustering result 1 and (c) Clustering result 2.

Step 2: Generate neuron n_j for sample i, then search each SGNT in the SGNF to find the neuron n_{win} at the shortest distance from neuron n_j.
Step 3: Connect the n_j into the SGNT covering the neuron n_{win}.
Step 4: Repeat Step 2 and Step 3 until all samples are input into the SGNF.

The generated SGNF includes c SGNTs, each SGNT corresponding to a cluster, and all the leaf neurons in a SGNT belong to the same cluster. Whereas, the number of clusters based on SGNF is user-specified.

The clustering results derived from the SGNF generating algorithm are influenced by the c seed samples used to generate the initial neuron forest. Suppose we partition the sample set X into c clusters $X_1, X_2, \cdots, X_c$; then their cluster centers are $m_1, m_2, \cdots, m_c$ respectively, and the between-class variance can be estimated

$$\sigma_B^2(X_1,X_2,\cdots,X_c) = \sum_{i=1}^{c-1}\sum_{j=i+1}^{c} p_i p_j (m_i - m_j)^2 \quad (1)$$

where $p_i = n_i/n$, n_i is the number of pixels for cluster X_i, and n is the total number of pixels. For an RGB image, $(m_i - m_j)^2$ is given by

$$(m_i - m_j)^2 = \sum_{q=1}^{3} (m_i^q - m_j^q)^2 \quad (2)$$

where m_i^q is the value of the ith cluster center in the qth color band.

According to the idea behind Otsu's thresholding method [22], the higher the between-class variance σ_B^2, the more accurate the partitioning of the sample set. Clustering with maximum σ_B^2 should subsequently yield an optimal partitioning of the sample set. In this context, the selection of the c seed samples can be modeled as an optimization problem.

Given n samples (pixels) in the data set, label them 1 to n, letting x_i; $i = 1,\ldots, c$ be the label numbers on the c seed samples. Then the x_i should satisfy two constraints

(1) $x_i \in \{1,\ldots, n\}$.
(2) $x_i \neq x_j$ if $i \neq j$.

There will be many collections of c samples that meet the above two constraints for a given sample set. Amongst these, the selection scheme having maximum σ_B^2 over the cluster partition obtained from the SGNF, is taken to be the optimal one.

As described next, by taking σ_B^2 as the fitness function, the GA is used to search for the c seed samples to optimize the clustering solution.

3.2. Genetic algorithm (GA)

GAs [23,24] are efficient and robust adaptive search techniques based on the idea of natural selection. The relevant steps of GA are:

Step 1: Randomly generate an initial population $G(0)$.
Step 2: Evaluate the fitness $f(m)$ of each individual m in the current population $G(t)$.
Step 3: Execute genetic operators including selection, crossover and mutation.
Step 4: Generate the next population $G(t+1)$ using genetic operators.
Step 5: Return to Step 2 until the maximum of the fitness function is obtained.

3.2.1. Chromosome coding and fitness function

Here a chromosome is coded by a c-integer string $chrom = (g_1, g_2, \ldots, g_k)$ where the g_i, $i = 1,\ldots, c$, are the genes taking the label numbers of the appropriate sub-regions as shown in Fig. 2(c). The c genes denote the selected seed samples with which the SGNF can be generated and the clustering image obtained. Following the constraints in Section 3.1, the values of the genes g_i, $i = 1,\ldots, c$ are different from each other.

The between-class variance $\sigma_B^2(X_1,X_2,\cdots,X_c)$ is used as a fitness function to evaluate the goodness of a chromosome. For a given chromosome, execute the SGNF generating algorithm using the c seed samples (genes) to obtain the image partition, then calculate the value of σ_B^2. The higher the value of σ_B^2, the better the chromosome quality is assumed to be. The chromosome having the maximum value of σ_B^2 is taken to be the optimal one in the population.

3.2.2. Genetic operators

The selection process copies individual strings into a tentative new population, called the mating pool, for genetic operations. The number of copies that an individual receives for the next generation is usually taken to be directly proportional to its fitness value, thereby mimicking the natural selection procedure. We utilize the "roulette wheel" selection approach.

The main purpose of crossover is to exchange information between randomly selected parent chromosomes by recombining parts of their genetic information. For simplicity, single-point crossover is used here. The probability of crossover p_c is set to fix the crossover rate, which is usually about 70%.

Mutation is the process by which a random alteration in the genetic structure of a chromosome takes place. Its main objective is to introduce genetic diversity into the population. Single-point mutation is used here. The probability of mutation p_m should usually be set fairly low. If it is set to high, the search will turn into a primitive random search.

The elitist strategy is applied after the genetic operations. The two "least fit" members of the new generation are replaced by the two "most fit" members in the current population. This guarantees that the fitness never declines from one generation to the next.

3.2.3. Termination criteria

We use very simple termination criteria: (i) when a plateau is reached whereby successive iterations no longer produce better results, and (ii) a fixed number of generations is reached.

3.3. Adaptive clustering based on GA and SGNN (ACluster-GA-SGNN)

When GA converges, c optimal seed samples are obtained by decoding the fittest individual in the population. The optimal clustering result is achieved by SGNF generated from the c seed samples. By example, selecting 2 samples to be seeds via GA, the clustering result for Fig. 2(a) using SGNN is shown in Fig. 4, where the samples are the sub-regions from the coarse segmentation in

Fig. 4. Clustering result for Fig. 2(a) based on SGNN and GA (the number of clusters is specified by the user).

Fig. 2(c), and the sample attributes consist of RGB colors and spatial location.

In the result depicted in Fig. 4, the number of clusters was user-specified. However, since such *a priori* information about the number of clusters in dermoscopy image is difficult to obtain, the clustering algorithms should be able to automatically determine this value. This problem can be addressed using indices of *cluster validity*. Cluster validity indices are commonly used for clustering evaluation and selection of optimal clustering schemes. A number of validity indices have been introduced, e.g., the Davies–Bouldin (DB) index [25], Dunn's index [26], and the SD index [27]. Most validity indices are based on two criteria: *Compactness*, where the idea is that the members of each cluster should be as close to each other as possible, and *Separation*, meaning, the clusters should be widely spaced. We utilize the SD index to determine the proper number of clusters.

The SD index is based on measurement of the scattering of clusters and the separation between clusters. Let $\sigma(X)$ be the variance of data set X and $\sigma(m_i)$ be the variance of cluster i. Then the scattering of clusters and the separation between clusters are

$$Scat(c) = \frac{1}{c}\sum_{i=1}^{c} \|\sigma(m_i)\| / \|\sigma(X)\| \tag{3}$$

$$Dis(c) = \frac{D_{max}}{D_{min}} \sum_{i=1}^{c}\sum_{j=1}^{c} \|m_i - m_j\|^{(-1)} \tag{4}$$

where c is the number of clusters, $D_{max} = \max(\|m_i - m_j\|)$ and $D_{min} = \min(\|m_i - m_j\|)$ are the maximum and minimum distances between cluster centers, respectively. The SD index is then

$$SD(c) = \alpha Scat(c) + Dis(c) \tag{5}$$

where $\alpha = Dis(c_{max})$ is a weighting factor, where c_{max} is the maximum number of input clusters.

Suppose the number of clusters for a given dermoscopic image lies between 2 and c_{max}, then calculate the SD index for every c using Eq. (5). Then the optimal number of clusters can be determined

$$c^* = \arg\min_{c} SD(c) \tag{6}$$

Using ACluster-GA-SGNN, the number of clusters can be determined adaptively, and the clustering task completed automatically. Fig. 5 shows an instance of adaptive clustering by ACluster-GA-SGNN on a malignant skin tumor image.

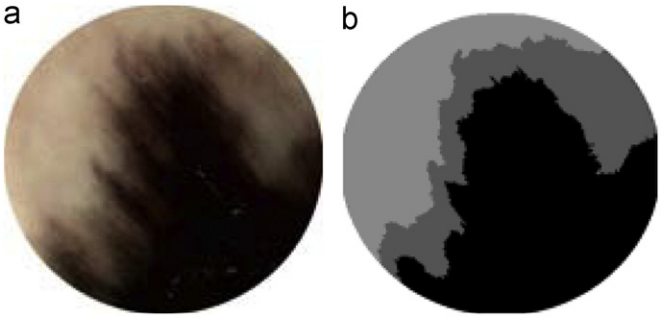

Fig. 5. Clustering results using ACluster-GA-SGNN (the number of clusters is automatic determined by the algorithm). (a) Original image and (b) Clustering result.

3.4. Automatic segmentation using GA and SGNN (ASegment-GA-SGNN)

The number of clusters varies across dermoscopy images, and the clustering image often includes more than two sub-regions that can be used for subsequent feature extraction and lesion segmentation. Features such as the number of clusters, the color and texture in each sub-region, and so on are important information for lesion classification. Actually, segmentation must be done before feature extraction, since the purpose is to determine those regions belonging to the lesion object and to detect the lesion border because firstly, the resulting border structure provides a basis for the calculation of important clinical features, such as lesion size and symmetry axes; and secondly, it is crucial for the extraction of some of the most discriminating dermoscopic features, such as radial streaming and pseudopods [8].

We place the focus on the segmentation of a lesion from its surrounding skin. Generally, the background skin region is brighter than the lesion region and will touch the image frame. Therefore, for an image with $c \geq 2$ clusters, the brightest cluster located at outer area in the image is taken as background skin, and the darkest cluster is as the lesion object firstly. And then, the rest clusters $X_i (1 \leq i \leq c-2)$ satisfying one of the following two items are merged into background skin.

(1) $I_{background} - I_i < k_I (I_i - I_{i-1})$, if $I_i < I_{background}$, where $I_i (I_i < I_{i+1})$ is the mean intensity of cluster X_i, and $I_{background}$ is the mean intensity of the first background skin cluster, and $0 < k_I < 1$ is a coefficient (we simply use $k_I = 1/2$).
(2) $touching\ Frame P_i > k_p border\ P_i$, if $|I_{background} - I_i| < 20$, where $touching\ Frame\ P_i$ is the number of pixels of cluster X_i touching the image frame, $border\ P_i$ is the number of border pixels of X_i, and $0 < k_p < 1$ is a coefficient (we simply use $k_p = 1/4$).

According to this merging schame, the clusters whose mean intensity is close to that of the brightest cluster and that have borders touching the image frame, will be merged into background skin. Fig. 6 is the segmentation result for Fig. 5(a), where Fig. 6(a) is the merging result for Fig. 5(b), and the red line in Fig. 6(b) is the border extracted from Fig. 6(a). This merging step is post-processing on the ACluster-GA-SGNN, so we call the proposed segmentation algorithm Segment-GA-SGNN.

4. Experimental results and analysis

We conducted a series of experiments using VC++6.0 on a Win7 OS with i5 3.1 GHz dual-core CPU and 2 GB RAM. There are two image datasets, one is of caucasians from [28], composed of 125 dermoscopy images including 68 malignant and 57 benign cases, with size from 227 × 252 to 512 × 768; the other one is of

xanthous race from the General Hospital of the Air Force of PLA of China, composed of 181 images including 53 malignant and 128 benign, with size from 560 × 560 to 560 × 752. We removed hair from the images using the method in [29]. Regarding the GA parameters, the population size is 25, the crossover probability p_c and the mutation probability p_m were taken to be 0.6 and 0.08 respectively, and the maximum number of generations is 40. Analysis is carried out with respect to the following three aspects:

4.1. Experiment 1: Clustering performance of ACluster-GA-SGNN

The ACluster-GA-SGNN model is compared against two widely used color clustering algorithms: *k*-means and FCM. For *k*-means and FCM, the number of clusters is specified by the user, and a filtering operator is carried out to remove noise. For the sake of fairness, the number of clusters for ACluster-GA-SGNN is specified by the user also. Fig. 7 is a group of clustering results, the first two rows are xanthous race images, and the last two rows are caucasian race images. It may be seen that when the contrast is large between the lesion object and its surrounding skin, and the lesion has a homogeneous texture and clear edge, then all three methods deliver a reasonable clustering result. When the object has no regular edges, shape or uniform color, and the contrast is reduced between the lesion object and its surrounding skin, then our model yields a more satisfying result.

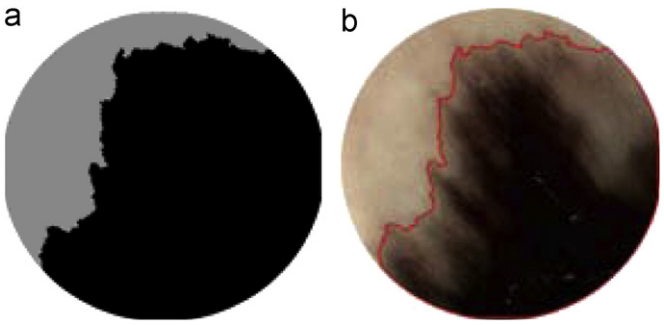

Fig. 6. Segmentation results using ASegment-GA-SGNN. (a) Segmentation result and (b) Extracted border.

The Jaccard and Minkowski scores are calculated in order to evaluate the accuracy of clustering [30,31]. Let T be the 'true' solution and C the solution a clustering algorithm generated. Let n_{11} be the number of pairs of data that are in the same cluster in both T and C. Let n_{10} be the number of pairs that are in the same cluster only in T, and n_{01} be the number of pairs that are in the same cluster only in C. Then the Jaccard score is defined as

$$S_J(T,C) = \frac{n_{11}}{n_{11}+n_{10}+n_{01}} \tag{7}$$

A higher value of $S_J(T,C)$ indicates a better clustering result; the Minkowski score is defined as

$$S_M(T,C) = \sqrt{\frac{n_{01}+n_{10}}{n_{11}+n_{10}}} \tag{8}$$

The smaller the $S_M(T,C)$ score, the better the solution is.

We computed the mean and deviation values of $S_J(T,C)$ and $S_M(T,C)$ on the caucasian and the xanthous datasets respectively. Table 1 states these results. From Table 1, FCM gives the lowest accuracy for both caucasian and xanthous dataset, *k*-means gives the best accuracy for the caucasian dataset, and ACluster-GA-SGNN gives the best clustering accuracy for the xanthous dataset. This follows since, when the number of clusters is specified, *k*-means can yield good accuracy, but when the image has weak contrast or variational color, ACluster-GA-SGNN can give more accurate results.

4.2. Experiment 2: Correctness of determining the number of clusters

Since it is difficult to obtain *a priori* information regarding the number of clusters in a dermoscopic image, in our model the SD validity index is used to automatically determine the number of clusters. The SD value can be calculated using Eq. (5).

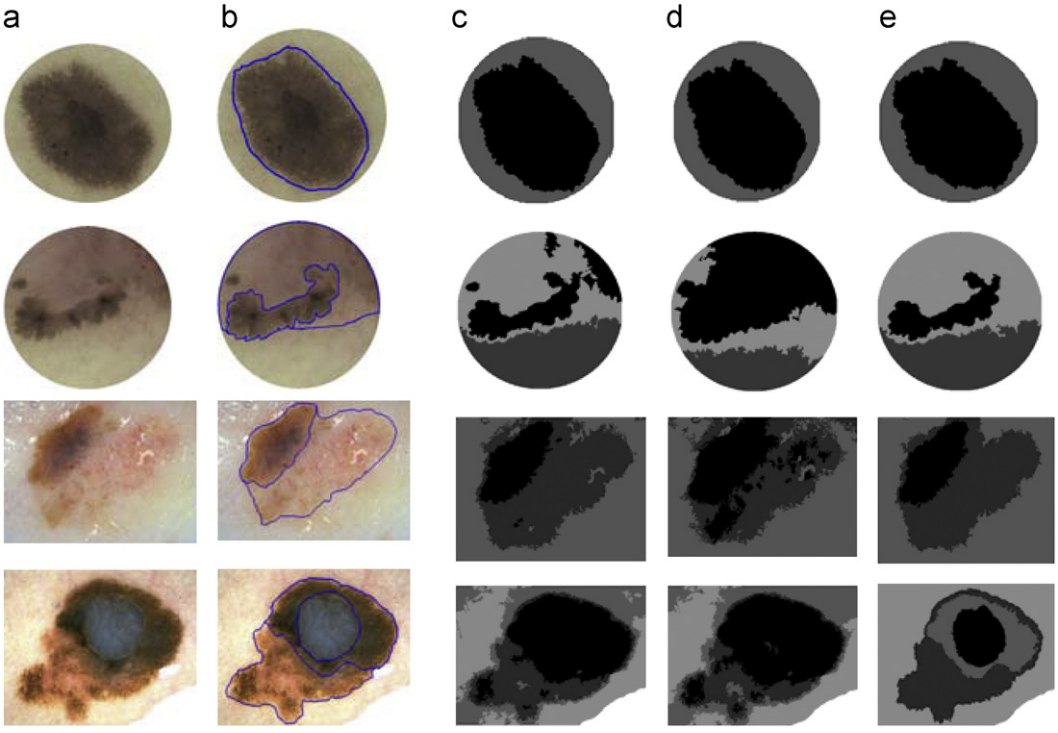

Fig. 7. Clustering results by three methods. (a) Original image, (b) Manual, border, (c) k-means, (d) FCM and (e) Our method.

Table 1
Statistics of clustering performance: mean (standard deviation).

Cluster method	*k*-means		FCM		ACluster-GA-SGNN	
Cluster measure	S_J	S_M	S_J	S_M	S_J	S_M
Caucasian race	0.834 (0.107)	0.410 (0.153)	0.776 (0.142)	0.493 (0.200)	0.826 (0.117)	0.425 (0.166)
Xanthous race	0.771 (0.168)	0.493 (0.279)	0.759 (0.194)	0.512 (0.262)	0.784 (0.152)	0.478 (0.278)

Table 2
Results of validity indices for the caucasian race dataset.

Actual number of clusters	Number of tested images	DB index	Dunn's index	SD index
2	66	66	66	57
3	57	2	5	35
4	2	0	0	1
Total	125	68	71	93

Table 3
Results of validity indices for the xanthous race dataset.

Actual number of clusters	Number of tested images	DB index	Dunn's index	SD index
2	108	102	102	90
3	71	5	6	47
4	2	0	0	2
Total	181	107	108	139

The minimum number of clusters is 2, and the maximum is 4. The weighting factor α, does not work well when set to $Dis(c_{max})$ as in [27], and is instead fixed at $\alpha = 3Dis(c_{max})$. Performance is analyzed using three validity indices, including the DB index, Dunn's index, and the SD index. For the tested image, the actual number of clusters is obtained visually as ground truth, and the results delivered by the validity indices is given in Tables 2 and 3.

Compared to DB and Dunn's indices, the SD index works well for 2–4 clusters, with correctness rate are 74.4% and 76.8% for the caucasian and the xanthous dataset respectively. Although the SD index is better suited for determining the number of clusters in dermoscopic images, it is clear that research on clustering validity is merited.

4.3. Experiment 3: Segment evaluation on ASegment-GA-SGNN

The otherwise convenient SGNN algorithm is sensitive to the order of input of the training samples, often yielding inconsistent results. GA is employed in our model to optimize SGNN clustering. In the post-processing phase, the segmentation is completed by merging the clustering regions into background skin and lesion objects. To evaluate segmentation performance, ASegment-GA-SGNNN is compared with the other automatic segmentation methods used on dermoscopic images, including Otsu's thresholding method, *k*-means, fuzzy *c*-means and SRM [7,32]. Fig. 8 shows a group of segmentation instances with the above algorithms, where the blue line is a manually inscribed border and the red line is the automatically determined border. The first two rows are xanthous race images and the last two rows are caucasian race images. It can be seen, when the lesion object has no uniform color, and weak contrast between the lesion and its surrounding skin, our method yields more accurate segmentation result. In Fig. 8, the third segmentation case for SRM (indicated by the blue line) is unsuccessful because of the weak contrast.

Segmentations were scored using the XOR metric [33], Hausdorff distance [34], and Jaccard similarity coefficient [31]. Let A be the ground truth and B the automatic segmentation result, then the XOR and Jaccard measures are given by

$$XOR = \frac{\text{Area}(A \cup B) - \text{Area}(A \cap B)}{\text{Area}(A)} \tag{9}$$

$$Jaccard = \frac{\text{Area}(A \cap B)}{\text{Area}(A \cup B)} \tag{10}$$

Area(I) denotes the number of pixels in the binary image I.

Let *borderA* and *borderB* be the border pixels of A and B respectively, then the Hausdorff distance is defined as

$$H(A,B) = \max(h(A,B), h(B,A)) \tag{11}$$

where

$$h(A,B) = \max_{a \in borderA} \min_{b \in borderB} \|a-b\| \tag{12}$$

and $\|\cdot\|$ is the Euclidean distance between two pixel points.

For XOR and Hausdorff distance, the lower the value, the better the segmentation result, whereas, a higher value of Jaccard similarity coefficient indicates a better segmentation result. Tables 4 and 5 give the results for the two datasets respectively. For SRM, there are two unsuccessful cases on the caucasian race dataset because of the weak contrast (see Fig. 8), and the statistical result for SRM in Table 4 excludes the unsuccessful cases. It can be seen that our method achieves the most accurate segmentation result among the automatic methods on both two datasets. The segmentation accuracy for the caucasian race dataset is better than the xanthous race dataset, probably because the images in the caucasian dataset have higher image quality than the xanthous dataset.

Computation time is also important for evaluating algorithms. For the two datasets, the computational time for the 5 segmentation methods was recorded in Table 6. Our model requires more time than the other methods, because of two reasons: GA needs a lot of time to search the resolution space, and the number of clusters needs to be determined. However, when the number of clusters is specified (as in *k*-means and FCM), our model only requires about 3.1 s for the caucasian and 4.0 s for the xanthous datasets.

5. Conclusion

We developed a model for the automatic segmentation of dermoscopic images based on a combination of SGNN and GA. SGNN is a competitive learning neural network which features simplicity in network design, fast learning and automatic organization capability. However, conventional SGNN is sensitive to the input order of the training samples, often delivering inconsistent cluster partitions. To optimize the clustering result GA is combined with SGNN in our model. SGNN is generalized from SGNT to SGNF, then a group of optimal seed samples is selected by GA. These seeds are used by SGNN to generate an optimal clustering partitioning of the dermoscopy images. The clustering regions can be used to analyze the color and texture of lesions; and they can

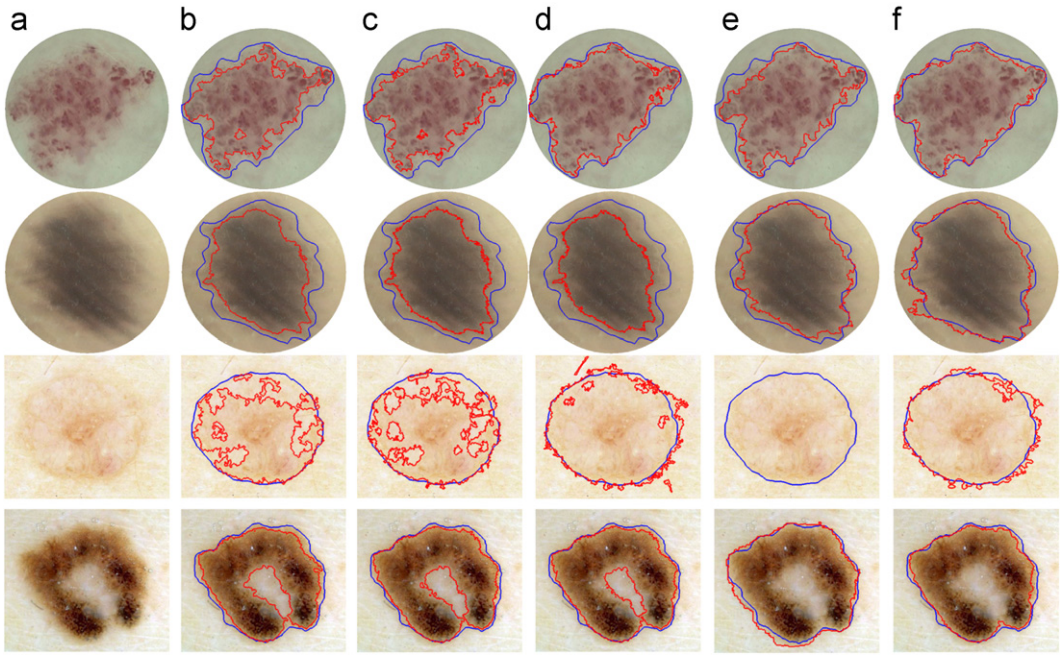

Fig. 8. Group of segmentation instances by automatic methods. The column from (a) to (f) is the original image, Otsu's threshold, *k*-means, FCM, SRM, and ASegment-GA-SGNN respectively.

Table 4
Statistics (%): mean(standard deviation) for the caucasian dataset.

Method	Otsu's threshold	*k*-means	FCM	SRM	Our method
XOR	20.3(13.2)	17.6(11.8)	19.4(13.0)	17.5(17.1)	15.0(9.6)
Hausdorff distance	49.9(42.4)	41.6(34.7)	53.2(41.5)	49.8(45.7)	43.0(41.2)
Jaccard index	79.8(13.2)	82.5(11.7)	80.9(12.8)	83.8(13.2)	85.5(9.2)

Table 5
Statistics (%): mean(standard deviation) for the xanthous dataset.

Method	Otsu's threshold	*k*-means	FCM	SRM	Our method
XOR	23.4(14.3)	23.3(14.1)	25.1(17.3)	23.7(19.4)	20.7(14.1)
Hausdorff distance	65.4(53.2)	65.1(53.6)	67.5(56.0)	74.3(57.8)	56.1(42.9)
Jaccard index	72.2(12.9)	77.5(12.6)	76.8(12.5)	79.6(14.0)	80.3(12.4)

Table 6
Computation time (s).

Method	Otsu's threshold	*k*-means	FCM	SRM	Our method
Caucasian race	0.3	0.8	1.3	0.2	9.3
Xanthous race	0.7 (0.6)	1.3	1.9	0.4	12.1

also be used to extract the lesion border, which is important for skin cancer diagnosis. The post-processing in our model entails merging the clustered regions into background skin and lesion objects, yielding a complete segmentation from which the border can be extracted. The overall ASegment-GA-SGNN model conducts an automatic and complete segmentation without manual interaction, and delivers more accurate segmentations as compared with other automatic methods.

Because of the complexity of skin lesions, it is difficult to obtain *a priori* information about the number of clusters in dermoscopy images. In the clustering phase, the SD validity index is used to automatically determine the number of clusters. Although the SD index is quite adaptable compared with two other popular indices, the error can still be large (25.6%for the caucasian and 23.2% for xanthous in our experiments). Clearly, more research is needed towards developing better cluster validity indices for dermoscopy images.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant nos. 61071138 and 61027004) and the Science Foundation of Beihang University(Grant no. YWF-12-LXGY-013). The authors thank Dr. Rusong Meng of the General Hospital of the Air Force of PLA at Beijing for providing the dermoscopy images.

References

[1] R. Siegel, E. Ward, O. Brawley, A. Jemal, Cancer statistics, 2011, CA: A Cancer Journal for Clinicians 61 (4) (2011) 212–236.
[2] F.R. Liu, Practical Skin Science, People Health Press, Beijing, 2005.
[3] G. Di Leo, C. Liguori, A. Paolillo, P. Sommella, An Improved Procedure for the Automatic Detection of Dermoscopy Structures in Digital ELM Images of Skin Lesions, in: Proceedings of the IEEE International Conference on Virtual

Environment, Human-Computer Interfaces, and Measurement Systems, Istanbul, Turkey, 2008.
[4] M. Binder, M. Chwarz, A. Winkler, A. Steiner, A. Kaider, K. Wolff, H. Pehamberger, Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesion for formally trained dermatologists, Archives of Dermatology 131 (1995) 286–291.
[5] J. Mayer, Systematic review of the diagnostic accuracy of dermoscopy in detecting malignant melanoma, Medical Journal of Australia 167 (1997) 206–210.
[6] H. Iyatomi, H.Oka, M.E. Celebi, M. Tanaka, K. Ogawa, Parameterization of Dermoscopic Findings for the Internet-Based Melanoma Screening System, in: Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing, Hawaii, 2007, pp. 189–193.
[7] M.E. Celebi, H. Kingravi, H. Iyatomi, A. Aslandogan, W.V. Stoecker, R.H. Moss, Border detection in dermoscopy images using statistical region merging, Skin Research and Technology 14 (3) (2008) 347–353.
[8] M.E. Celebi, H. Iyatomi, G. Schaefer, W.V. Stoecker, Lesion border detection in dermoscopy images, Computerized Medical Imaging and Graphics 33 (2) (2009) 148–153.
[9] C. Grana, G. Pellacani, R. Cucchiara, S. Seidenari, A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions, IEEE Transactions on Medical Imaging 22 (8) (2003) 959–964.
[10] P. Rubegni, A. Ferrari, G. Cevenini, D. Piccolo, M. Burron, et al., Differentiation between pigmented spitz naevus and melanoma by digital dermoscopy and stepwise logistic discriminant analysis, Melanoma Research 11 (1) (2001) 37–44.
[11] H. Zhou, G. Schaefer, M.E. Celebi, F. Lin, T. Liu, Gradient vector flow with mean shift for skin lesion segmentation, Computerized Medical Imaging and Graphics 35 (2) (2011) 121–127.
[12] M.E. Celebi, A. Aslandogan, W.V. Stoecker, Unsupervised border detection in dermoscopy images, Skin Research and Technology 13 (4) (2007) 454–462.
[13] H. Zhou, G. Schaefer, A.H. Sadka, M.E. Celebi, Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images, IEEE Journal of Selected Topics in Signal Processing 3 (1) (2009) 26–34.
[14] J. Gao, J. Zhang, M.G. Fleming, A Novel Multiresolution Color Image Segmentation Technique and its Application to Dermatoscopic Image Segmentation, in: Proceedings of the IEEE International Conference on Image Processing, Vancouver, BC, Canada, September 2000.
[15] R. Melli, C. Grana, R. Cucchiara, Comparison of Color Clustering Algorithms for Segmentation of Dermatological Images, in: Proceedings of the SPIE Conference on Medical Imaging, San Diego, vol. 3S1–9, 2006.
[16] M. Silveira, J.C. Nascimento, J.S. Marques, A.R.S. Marcal, T. Mendonca, S. Yamauchi, J. Maeda, J. Rozeira, Comparison of segmentation methods for melanoma diagnosis in dermoscopy images, IEEE Journal of Selected Topics in Signal Processing 3 (1) (2009) 35–45.
[17] R. Cucchiara, C. Grana, S. Seidenari, G. Pellacani, Exploiting color and topological features for region segmentation with recursive fuzzy C-means, Machine Graphics and Vision 11 (2/3) (2002) 169–182.
[18] W.X. Wen, A. Jennings, H. Liu, Learning a Neural Tree, in: Proceedings of the International Journal Conference on Neural Networks, 1992, pp. 751–756.
[19] H. Inoue, H. Narihisa, Efficiency of self-generating neural networks applied to pattern recognition, Mathematical and Computer Modelling 38 (2003) 1225–1232.
[20] H. Inoue, H. Narihisa, Efficient pruning method for ensemble self-generating neural networks, Journal of Systemics, Cybernetics and Informatics 1 (6) (2003) 72–77.
[21] S. Feng, A. Tan, Self-Organizing Neural Networks for Behavior Modeling in Games, in: Proceedings of the International Journal Conference on Neural Network, Barcelona, 2010, pp. 1–8.
[22] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man and Cybernetics 9 (1) (1979) 62–66.
[23] E. Zacharia, D. Maroulis, An original genetic approach to the fully automatic gridding of microarray images, IEEE Transactions on Medical Imaging 27 (6) (2008) 805–813.
[24] U. Maulik, S. Bandyopadhyay, Genetic algorithm based clustering technique, Pattern Recognition 33 (2000) 1455–1465.
[25] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence 1 (1979) 224–227.
[26] J.C. Dunn, Well separated clusters and optimal fuzzy partitions, Journal of Cybernetics and Systems 4 (1974) 95–104.
[27] M. Halkidi, M. Vazirgiannis, Y. Batistakis, Quality Scheme Assessment in the Clustering Process, in: Proceedings of the European Conference on Principles and Practice of Knowledge Discovery in Databases, Lyon, France, 2000, pp. 265–276.
[28] H. Iyatomi, H. Oka, M.E. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, K. Ogawa, An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm, Computerized Medical Imaging and Graphics 32 (7) (2008) 566–579.
[29] F.Y. Xie, S.Y. Qin, Z.G. Jiang, R.S. Meng, PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma, Computerized Medical Imaging and Graphics 33 (4) (2009) 275–282.
[30] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithm, Plenum, New York, 1981.
[31] D.W. Kim, K.H. Lee, D. Lee, Detecting clusters of different geometrical shapes in microarray gene expression data, Bioinformatics 21 (9) (2005) 1927–1934.
[32] R. Nock, F. Nielsen, Statistical region merging, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (11) (2004) 1452–1458.
[33] M.E. Celebi, G. Schaefer, H. Iyatomi, W.V. Stoecker, J.M. Malters, J.M. Grichnik, An improved objective evaluation measure for border detection in dermoscopy images, Skin Research and Technology 15 (4) (2009) 444–450.
[34] D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge., Comparing images using the Hausdorff distance, IEEE Transaction on Pattern Analysis and Machine Intelligence 15 (9) (1993) 850–863.

Fengying Xie received the B.Eng. degree from the Department of Computer Science, Daqing Petroleum Institute, Daqing, China, in 1997, the M.S. and Ph.D. Degree in Pattern Recognition and Intelligent System from Beihang University, Beijing, China, in 2002 and 2009 respectively. She was a visiting scholar in the Laboratory for Image and Video Engineering (LIVE) at the University of Texas at Austin from 2010 to 2011. She is now an associate professor at the School of Astronautics in BeiHang University. Her research interests include biomedical image processing, image analysis and segmentation, object detection and recognition.

Alan Conrad Bovik received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1980, 1982, and 1984, respectively.

He is the Curry/Cullen Trust Endowed Chair Professor at The University of Texas at Austin, where he is the Director of the Laboratory for Image and Video Engineering (LIVE). He is a faculty member in the Department of Electrical and Computer Engineering and the Center for Perceptual Systems in the Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has published over 500 technical articles in these areas and holds two U.S. patents. He is the author of The Handbook of Image and Video Processing (Academic Press, 2005), Modern Image Quality Assessment (Morgan & Claypool, 2006), and two new books, The Essential Guide to Image Processing and The Essential Guide to Video Processing (Academic Press).

Dr. Bovik has received a number of major awards from the IEEE Signal Processing Society, including: the Best Paper Award (2009); the Education Award (2007); the Technical Achievement Award (2005); Distinguished Lecturer (2000); and the Meritorious Service Award (1998). He also received the Hocott Award for Distinguished Engineering Research at the University of Texas at Austin (2008); the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana (2008); the IEEE Third Millennium Medal (2000); and two journal paper awards from the international Pattern Recognition Society (1988 and 1993). He is a Fellow of the IEEE, a Fellow of the Optical Society of America, and a Fellow of the Society of Photo-Optical and Instrumentation Engineers. He has been involved in numerous professional society activities, including: Board of Governors, IEEE Signal Processing Society, 1996–1998; Editor-in-Chief, IEEE Transactions on Image Processing, 1996–2002; Editorial Board, The Proceedings of the IEEE, 1998–2004; Series Editor for Image, Video, and Multimedia Processing, Morgan and Claypool Publishing Company, 2003-present; and Founding General Chairman, First IEEE International Conference on Image Processing, held in Austin, Texas, in November, 1994.

Dr. Bovik is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial and academic institutions.