Automated Basal Cell Carcinoma Detection in High-Definition Optical Coherence Tomography

Annan Li, Jun Cheng, Ai Ping Yow, Ruchir Srivastava, Damon Wing Kee Wong,
Hong Liang Tey and Jiang Liu

***Abstract*— Basal cell carcinoma (BCC) is the most common non-melanoma skin cancer. Conventional diagnosis of BCC requires invasive biopsies. Recently, a high-definition optical coherence tomography (HD-OCT) technique has been developed, which provides a non-invasive *in vivo* imaging method of skin. Good agreements of BCC features between HD-OCT images and histopathological architecture have been found. Therefore it is possible to automatically detect BCC using HD-OCT. This paper presents a novel BCC detection method that consists of four steps: graph based skin surface segmentation, surface flattening, deep feature extraction and the BCC classification. The effectiveness of the proposed method is well demonstrated on a dataset of 5,040 images. It can therefore serve as an automatic tool for screening BCC.**

I. Introduction

Basal cell carcinoma (BCC) is the most common malignant skin tumor, which accounts for approximately 70% of non-melanoma skin cancers [1]. Conventional diagnostic procedure requires a histopathological examination based on tissue biopsy of the tumour, which is invasive and implicates multiple surgical procedures. In the past decade, non-invasive *in vivo* imaging techniques have been introduced to the detection of BCC, including high-frequency ultrasound (HFUS), optical coherence tomography (OCT) and reflectance confocal microscopy (RCM). The penetration of HFUS is the deepest but its resolution is very limited, while RCM has the highest resolution but the shallowest penetration. The resolutions an the penetration depth of conventional OCT both lie between HFUS and RCM.

It has been reported that the specificity of HFUS and conventional OCT in the diagnosis of BCC is insufficient due to low resolution [2]. Although RCM provides a clearer image, its penetration depth is very limit. Recently, a new high definition optical coherence tomography (HD-OCT) device has been developed [3], which provides a resolution comparative to RCM and a penetration depth close to HFUS. Recent studies [4], [2] show good agreements of BCC features between HD-OCT images and histological sections, which implies that it is possible to automatically detect BCC using HD-OCT.

In this work, a novel approach for automatic BCC detection is proposed. First, the skin surface is extracted by a graph based method. Then the skin image is normalized by flattening the surface. BCC detection is achieved by learning a classifier on the normalized images. Deep convolutional neural networks have been prove successful in image classification. It provides an end-to-end solution for image classification. However, training an effective network requires a large number of images, which is not available in BCC detection. We propose an alternative approach by adopting pre-trained networks that obtained from large-scale dataset and integrating it with the skin images via a discriminant classifier. The effectiveness of proposed method is well demonstrated on a dataset consists of 5,040 images. To the best of our knowledge, it is the first study on automatic BCC detection in HD-OCT.

The remainder of this paper is organized as follows. Section II describes the proposed BCC detection method. Section III shows the experimental results. Conclusions and discussions are given in Section IV.

II. Methodology

As shown in Figure 1, the proposed BCC detection method includes four main steps: skin surface extraction, surface flattening, deep feature extraction and BCC classification.

A. Skin Surface Segmentation

The HD-OCT image is captured by placing the probe onto target skin area. Optic gels are applied for enhancing the signal, which result in the dark area between the skin surface and the top of the image. Due to the difference in gel thickness and the pressure of placing the probe, the vertical locations of skin surface varies in image (see Figure 2). Besides that, the complicated topography of skin surface also makes it difficult to directly compare the image. Therefore, image normalization is a prerequisite for detecting BCC. Consequently, skin surface segmentation is needed.

We represent a skin image as an undirected graph, in which each pixel is a vertex. Having the prior knowledge that skin surface is a continuous contour across the entire width of image, the connectivity of the contour should be enforced. It is achieved by defining the edges of graph as 8-neighbor connections of every pixel. Consequently, the skin surface segmentation can be formulated as a problem of finding the

This work was supported in part by the Agency for Science, Technology and Research, Singapore, under A*STAR-NHG-NTU Skin Research Grant.

A. Li, J. Cheng, A. Yow, R. Srivastava, and D.W.K. Wong are with the Institute for Infocomm Research, Agency for Science, Technology and Research, 138632, Singapore. (E-mail: {lia, jcheng, yowap, srivastava, wkwong}@i2r.a-star.edu.sg)

H. Tey is with National Skin Centre, 308205, Singapore. (E-mail: hltey@nsc.gov.sg)

J. Liu is with Ningbo Institute of Industrial Technology, Chinese Academy of Sciences. 1219 Zhongguan West Road, Zhenhai, Ningbo, Zhejiang 315201, China. (E-mail: jimmyliu@nimte.ac.cn)

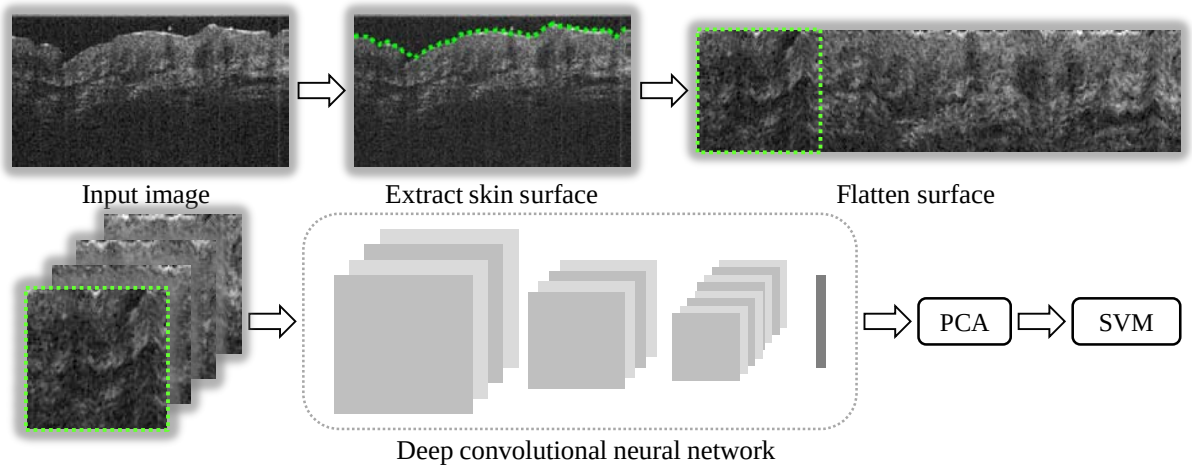

Fig. 1: Pipeline of BCC detection in HD-OCT.

shortest path of the graph [5]. The weight w of a graph edge of is defined as

$$w(m,n) = (1-G_m) + (1-G_n) + \sigma, \tag{1}$$

where $0 \leq G_m \leq 1$ and $0 \leq G_n \leq 1$ are normalized vertical gradients for pixel m and n. $\sigma > 0$ is a regularization constant of small value. The solution of the shortest path is achieved by Dijkstra's algorithm [5].

B. Image Normalization

The skin image is normalized by flattening the surface. Denote the skin surface as a set of vertical coordinates $Y = [Y_1, Y_2, \ldots, Y_n]$, where n is the width of the image. The flattened image is simply obtained by

$$Im^{flatte}(x,y) = Im^{original}(x, y - Y_x), \tag{2}$$

where x is the column index. By fixing the sampling depth to 90 pixel, the skin area is flattened to a 360×90 rectangle. Exemplar images of flattened skin can be found in Figure 2.

C. Deep Feature Extraction

Convolutional networks (ConvNets) have been proved better than handcrafted features in large-scale image classification. As mentioned in Section I, due to the relatively limit number of skin images, ConvNets is only used for feature representation. 4 popular ConvNets obtained from the ImageNet [6] dataset are adopted to BCC detection: *AlexNet* [7], *VGG-16*, *VGG-19* [8] and *GoogLeNet* [9].

The architecture of a ConvNet consists of a stack of convolutional layers followed by several fully connected layers. We find that the fully connected layers in AlexNet, VGG-16 and VGG-19 do not enhance the performance. Only the convolutional layers are used in this work. The configuration of convolutional layers is determined by three factors: the filter size, the channel or filter number and the followed max pooling layer. Table I shows the configurations of AlexNet, VGG-16 and VGG-19. As can be seen, AlexNet has bigger filter sizes but fewer convolutional layers, while VGG-16 and VGG-19 are smaller in filter size. The key difference between GoogLeNet and AlexNet, VGG-16 and VGG-19 is the introduction of additional 1×1 convolutional layers and the *Network in Network (NIN)* inspired by Lin et al. [10]. NIN is a small branching and merging network structure embedded in convolutional networks. By applying this special architecture, a much deeper but computational efficient network that consists of 27 layers is developed.

TABLE I: Configurations of AlexNet, VGG-16 and VGG-19.

AlexNet		VGG-16		VGG-19	
Input (Image Patch)					
filter	channel	filter	channel	filter	channel
11×11	96	3×3 3×3	64 64	3×3 3×3	64 64
Max pooling					
5×5	256	3×3 3×3	128 128	3×3 3×3	128 128
		Max pooling			
		3×3 3×3 3×3	256 256 256	3×3 3×3 3×3 3×3	256 256 256 256
Max pooling					
3×3	384	3×3 3×3 3×3	512 512 512	3×3 3×3 3×3 3×3	512 512 512 512
3×3	384	Max pooling			
3×3	256	3×3 3×3 3×3	512 512 512	3×3 3×3 3×3 3×3	512 512 512 512
Max pooling					
Output (4096 dimensional feature)					

As shown in Figure 1, the shape of the normalized image is flat which is different from the square template used in ImageNet classification [7], [8], [9]. To apply the ConvNets, the HD-OCT image is divided into 4 90×90 non-overlapping square patches, from which deep features are extracted independently.

The dimensions of output feature for AlexNet, VGG-16,

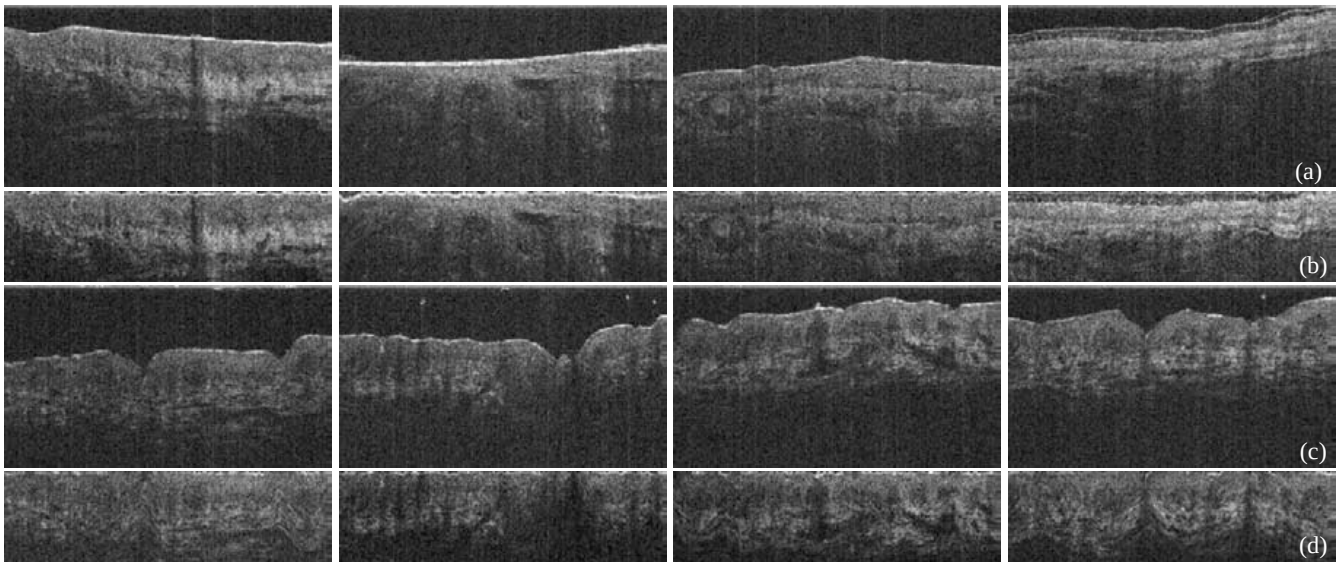

Fig. 2: Exemplar HD-OCT images of skin. Row (a) and (b) show some diseased images before and after flattening respectively, while (c) and (d) are normal cases.

VGG-19 and GoogLeNet are 4,096, 4,096, 4,096 and 1000 respectively. Concatenating the features of 4 patches, a HD-OCT image is represented by a 16,384 or 4,000 dimensional vector. As can be seen, the dimension is very high which may lead to heavy computational cost. In practice, the BCC detection will be performed in the whole HD-OCT volume that consists of hundreds of images rather than a single image. The diagnosis will be based on the number of abnormal image. Therefore computational efficiency is important. To speed up the detection, principal component analysis (PCA) [12] is applied for dimension reduction.

D. Basal Cell Carcinoma Detection

The detection of BCC is formulated as a binary classification problem based on the deep features described in Section II-C. In this work, support vector machine (SVM) [11] is used to classify the diseased images from normal cases.

III. Experiments

The proposed method is evaluated on a dataset consists of 14 three dimensional HD-OCT volumes, including 7 BCC and 7 normal cases. All the samples are captured by a SkinTell® [3] HD-OCT machine. The volume size is $360 \times 200 \times 360$ that corresponds to a $1.0 \times 0.57 \times 1.0$ mm cuboid. Consequently there are 5,040 360×200 B-scans, in which 1,875 contains lesion or irregular structures. All the disease cases are nodular BCC on the human face. To make a fair comparison, the normal samples on corresponding area are also included. Examples of the HD-OCT images can be found in Figure 2.

Considering that adjacent images in the same HD-OCT volume are highly correlated, the results might be misleading if such images appear in both training and testing set. The experiments are carried out by performing a 7-fold cross validation. In each fold, 1 BCC and 1 normal volumes are used for testing, the rest 12 volumes are used for training the SVMs. With such an arrangement, the influence of adjacent image is strictly excluded in evaluation. The implementation of classifier is based on the LIBSVM library [11]. Linear SVMs with default parameters are used for detecting the BCC.

BCC detection results in the form of receiver operating characteristic (ROC) curves are shown in Figure 3. Corresponding area under the curves (AUCs) are given in Table II. The performance rank is VGG-16>AlexNet>VGG-19>GoogLeNet using full length of features. The highest AUC is 0.935, which clearly proves that automatic BCC detection in HD-OCT is a feasible approach.

Experimental results with PCA dimension reduction are also presented. As can be seen, the PCA modeling can considerably reduce the feature dimension with a small sacrifice of performance. AlexNet and VGG-19 even get better results after dimension reduction. Among the 4 ConvNets, GoogLeNet gets no performance decline even when the feature dimension is reduced to 100. Because of the limit dimension of original features, GoogLeNet does not achieve as good performance as AlexNet, VGG-16 and VGG-19. However, the experimental results show that feature of GoogLeNet is a very stable. AlexNet achieves the second smallest performance decline. When the feature dimension is reduced to 100, AlexNet outperforms other ConvNets and obtain a AUC loss of 0.019. VGG-16 would be the choice of ConvNet, however, if the speed requirement is high AlexNet with PCA compression is better.

IV. Conclusions and Discussions

In this paper, we first study the automatic detection of BCC in HD-OCT. A novel BCC detection approach is proposed, which includes four steps: graph based skin surface extrac-

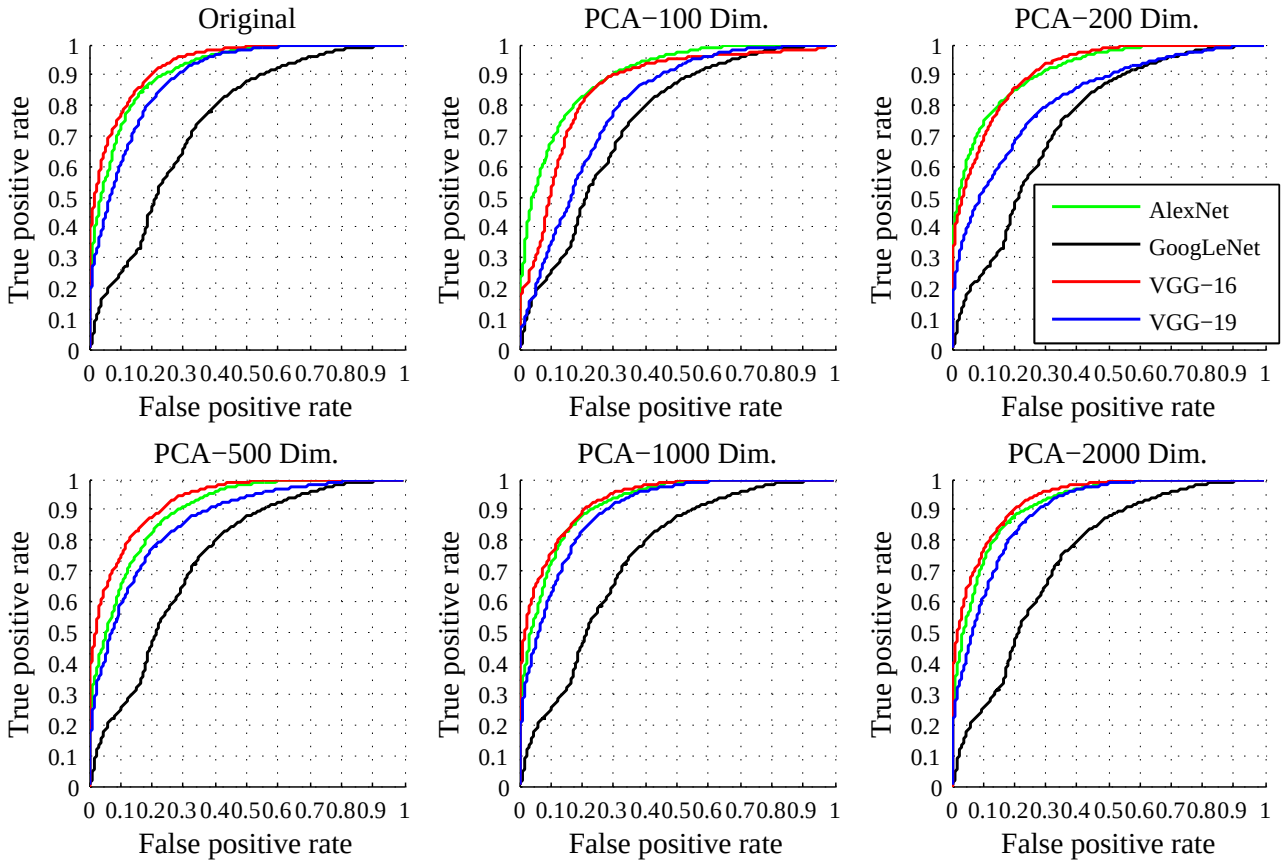

Fig. 3: Basal cell carcinoma detection results in receiver operating characteristic curves.

TABLE II: Area under the curves (AUCs).

ConvNet	Ori.	Dimension of PCA feature				
		100	200	500	1000	2000
AlexNet	0.916	0.897	0.915	0.897	**0.917**	0.917
GoogLeNet	0.744	0.744	0.744	0.744	0.744	0.744
VGG-16	**0.935**	0.858	0.913	0.928	0.931	0.933
VGG-19	0.891	0.798	0.824	0.863	**0.894**	0.893

tion, surface flattening, deep convolutional feature extraction and SVM based BCC classification. Experimental validation on a dataset consists of 5,040 images clearly shows that the proposed method is effective and using HD-OCT to detect BCC is a feasible and promising approach.

Training an effective ConvNet usually requires millions images. Severe overfitting will occur if the sample number is insufficient. Although large-scale HD-OCT dataset of skin is currently not available, fine tuning the pre-trained ConvNets on skin images might improve the performance. Exploring the fine tuning of ConvNet will be our future work.

REFERENCES

[1] C S Wong, R C Strange, and J T Lear. Basal cell carcinoma. *British Medical Journal*, 327(7418):794, 2003.

[2] T Maier, M Braun-Falco, T Hinz, MH Schmid-Wendtner, T Ruzicka, and C Berking. Morphology of basal cell carcinoma in high definition optical coherence ttomography: En-face and slice imaging mode, and comparison with histology. *Journal of the European Academy of Dermatology and Venereology*, 27(1):e97–e104, 2013.

[3] Agfa-HealthCare: Technical Specifications of SKINTELL. http://www.agfahealthcare.com/

[4] Marc ALM Boone, Sarah Norrenberg, GBE Jemec, and Véronique Del Marmol. Imaging of Basal Cell Carcinoma by High-Definition Optical Coherence Tomography: Histomorphological Correlation. A Pilot Study. *British Journal of Dermatology*, 167(4):856–864, 2012.

[5] Annan Li, Jun Cheng, Ai Ping Yow, Carolin Wall, Damon Wing Kee Wong, Hong Liang Tey, and Jiang Liu. Epidermal Segmentation in High-Definition Optical Coherence Tomography. In *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3045–3048. IEEE, 2015.

[6] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[8] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[9] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going Deeper with Convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[10] Min Lin, Qiang Chen, and Shuicheng Yan. Network in Network. *arXiv preprint arXiv:1312.4400*, 2013.

[11] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27, 2011.

[12] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2001.