Computer Methods and Programs in Biomedicine 162 (2018) 221–231

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks

Mohammed A. Al-masni[a,1], Mugahed A. Al-antari[a,1], Mun-Taek Choi[b], Seung-Moo Han[a], Tae-Seong Kim[a,*]

[a] Department of Biomedical Engineering, College of Electronics and Information, Kyung Hee University, Yongin, Republic of Korea
[b] School of Mechanical Engineering, Sungkyunkwan University, Republic of Korea

ARTICLE INFO

Article history:
Received 20 March 2018
Revised 30 April 2018
Accepted 17 May 2018

Keywords:
Deep learning
Dermoscopy
Full resolution convolutional network (FrCN)
Melanoma
Skin lesion segmentation

ABSTRACT

Background and objective: Automatic segmentation of skin lesions in dermoscopy images is still a challenging task due to the large shape variations and indistinct boundaries of the lesions. Accurate segmentation of skin lesions is a key prerequisite step for any computer-aided diagnostic system to recognize skin melanoma.

Methods: In this paper, we propose a novel segmentation methodology via full resolution convolutional networks (FrCN). The proposed FrCN method directly learns the full resolution features of each individual pixel of the input data without the need for pre- or post-processing operations such as artifact removal, low contrast adjustment, or further enhancement of the segmented skin lesion boundaries. We evaluated the proposed method using two publicly available databases, the IEEE International Symposium on Biomedical Imaging (ISBI) 2017 Challenge and PH2 datasets. To evaluate the proposed method, we compared the segmentation performance with the latest deep learning segmentation approaches such as the fully convolutional network (FCN), U-Net, and SegNet.

Results: Our results showed that the proposed FrCN method segmented the skin lesions with an average Jaccard index of 77.11% and an overall segmentation accuracy of 94.03% for the ISBI 2017 test dataset and 84.79% and 95.08%, respectively, for the PH2 dataset. In comparison to FCN, U-Net, and SegNet, the proposed FrCN outperformed them by 4.94%, 15.47%, and 7.48% for the Jaccard index and 1.31%, 3.89%, and 2.27% for the segmentation accuracy, respectively. Furthermore, the proposed FrCN achieved a segmentation accuracy of 95.62% for some representative clinical benign cases, 90.78% for the melanoma cases, and 91.29% for the seborrheic keratosis cases in the ISBI 2017 test dataset, exhibiting better performance than those of FCN, U-Net, and SegNet.

Conclusions: We conclude that using the full spatial resolutions of the input image could enable to learn better specific and prominent features, leading to an improvement in the segmentation performance.

1. Introduction

Melanoma skin cancer is one of the most common cancers in the world. It originates when a certain type of skin cell, called a melanocyte, begins to grow out of control. This kind of skin cancer is considered a malignant tumor and named as melanoma [1]. According to the annual report of the American Cancer Society in the United States [2], about 87,110 cases were diagnosed as new cases of melanoma, and the estimated deaths from this disease included up to 9730 cases in 2017. It is reported that melanoma is the most deadly skin cancer, with a mortality rate of 1.62% among all other cancers [3]. It is extremely important to diagnose melanoma in its early stage, since this allows treatments and increases the survival rate [4]. Visual inspection during the medical examination of skin lesions suffers from the similarity among the skin lesions and normal tissues, which may result in inaccurate diagnoses [5]. In the past decade, dermoscopy has been used to aid dermatologists in improving the screening for melanoma skin lesions. Dermoscopy is a non-invasive imaging tool that acquires a magnified image of the skin lesion utilizing polarized light [6]. It visualizes deeper details of the skin structure by abolishing the reflection of the skin surface. Although dermoscopy improves the diagnostic accuracy compared to that obtained via visual inspection, screening through dermoscopy images by dermatologists is still complex, time consum-

* Corresponding author.
E-mail addresses: m.almasani@khu.ac.kr (M.A. Al-masni), en.mualshz@khu.ac.kr (M.A. Al-antari), mtchoi@skku.edu (M.-T. Choi), smhan@khu.ac.kr (S.-M. Han), tskim@khu.ac.kr (T.-S. Kim).
[1] Authors equally contributed to this work.

https://doi.org/10.1016/j.cmpb.2018.05.027
0169-2607/

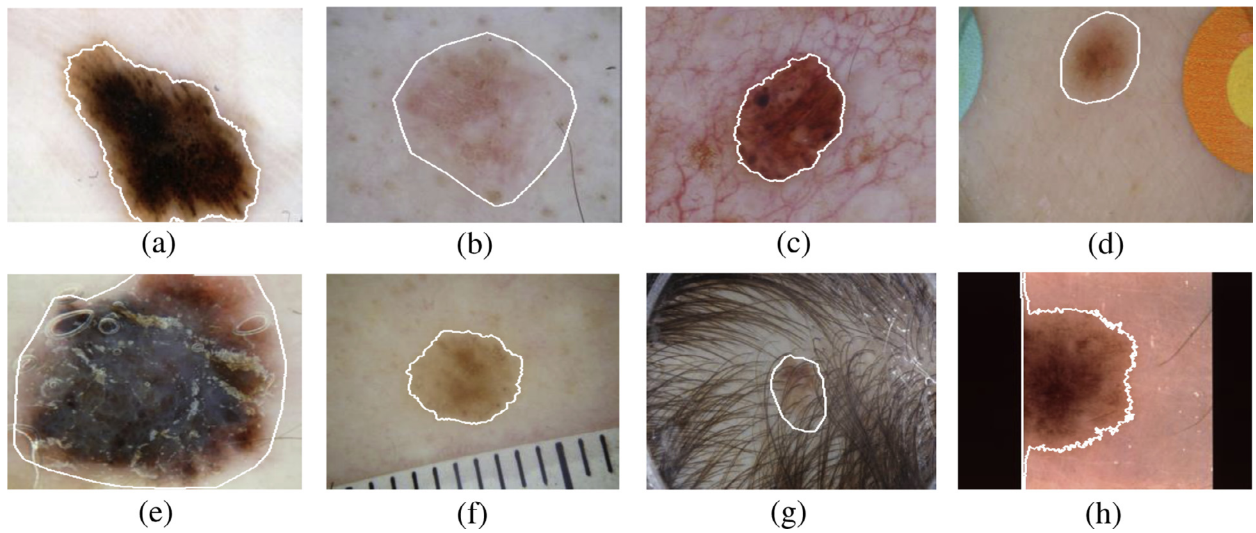

Fig. 1. Examples of some challenging cases of skin lesions such as (a) irregular fuzzy boundaries, (b) low contrast, (c) blood vessels, (d) color illumination, (e) bubbles, (f) ruler mark artifact, (g) hair artifact, and (h) frame artifact. White contours indicate the lesions segmented by expert dermatologists. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ing, subjective, and fault prone [5]. Hence, automated computerized diagnostic systems for skin lesions are highly necessary to help and support dermatologists in decision-making.

An essential preliminary step in any computerized diagnostic system of melanoma is to segment the skin lesions automatically [7,8,9]. Accurate segmentation of the skin lesions from surrounding tissues plays a crucial role in acquiring more specific and prominent features, which are utilized to classify the melanoma. In fact, it is essential to exclude normal tissues via segmentation and extract more representative features from the lesions for accurate diagnosis [9,10]. However, such segmentation is a challenging task due to the large variations among skin lesions in shapes, size, color, texture, and location in the dermoscopy images. Also, their low contrast from neighboring tissues poses additional difficulties. Furthermore, extra factors such as hair, blood vessels, ruler marks, air bubbles, ebony frames, and color illumination contribute extra obstacles to the segmentation task. Fig. 1 illustrates some examples of these challenges.

In general, segmentation techniques for skin cancer lesions can be categorized into the following five groups. *Histogram thresholding methods* determine one or more proper threshold values to differentiate the skin lesions from the surrounding tissues [11,12,13,14,15]. *Unsupervised clustering methods* utilize the color space features of RGB dermoscopy images to acquire homogenous regions [16,17,18,19]. *Edge-based* and *region-based methods* utilize the edge operator and region splitting or merging algorithms, respectively [20,21,11,22]. Active contour methods (e.g., *snakes active contours*) adapt the evolution curve algorithm to segment skin lesions [11,23,24,25]. Lastly, *supervised segmentation methods* segment the skin lesions by training the recognizers such as decision trees (DTs), support vector machines (SVMs), and artificial neural networks (ANN) [16,26]. For further details of these methods, the latest comprehensive reviews of the segmentation methods for skin lesions provide more details [7,8,27,28]. In fact, all of these techniques used the low level attributes that rely only on pixel level features. Thus, these conventional segmentation methods still do not provide satisfactory performance and cannot overcome the challenges of low contrast and hair artifacts.

Recently, segmentation techniques employing deep learning approaches have been getting major attention in the fields of object localization, semantic segmentation, and image classification [29,30,31,32,33,34,35]. Deep learning based on a convolutional neural network (CNN) is a powerful technique since it has the capability to extract more prominent features from the entire image rather than hand-crafted features [36,37,38]. In 2013, the initial CNN approach to the segmentation task was introduced by [39,40]. They applied a pixel segmentation method by dividing the input image into smaller patches (i.e., sliding windows) and passing them into a CNN classifier. The prediction of each patch represented the segmentation of the center pixel in the corresponding patch. They performed this segmentation method only for some rib regions in the chest X-ray images. In 2015, Melinščak et al. proposed a similar CNN segmentation approach for retinal vessels as a binary classification task [41]. The prediction (i.e., vessel or non-vessel) for each patch determined the segmentation of its central pixel. This approach still needs further improvement since the deep attributes extracted from each patch do not exactly reflect the center pixel features. Also, its processing requires a lot of time since the network should be executed separately for each patch [42].

Lately, state-of-the-art approaches have been developed using deep CNN methods to enhance the performance of challenging segmentation tasks. In 2015, Long et al. developed the first end-to-end pixel-wise semantic segmentation technique called a fully convolutional network (FCN) [43,44]. They adapted the well-known recognition models of AlexNet, GoogLeNet, and Visual Geometry Group (VGG)Net into FCNs by substituting the fully connected layers with the convolutional layers. They upsampled the last convolutional layer with the deconvolution and bilinear interpolation processes to produce a map that involves dense predictions with the same size as the input image. FCN achieved a significant improvement in segmentation accuracy on the Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes (PASCAL VOC) dataset against the conventional methods [45]. In 2015, Noh et al. developed a deep learning segmentation method called a deconvolution network (DeconvNet) as an extension of FCN [46]. DeconvNet consisted of a convolution network to extract the features from input images and a deconvolution network to generate object segmentation. DeconvNet utilized the VGGNet of 16 layers in the convolution network, and the same layers were mirrored in the deconvolution network. This method produced a dense map that had a higher spatial resolution compared to that generated by FCN. In 2015, Badrinarayanan et al. proposed a deep convolutional encoder-decoder segmentation method