2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Skin Lesion Segmentation in Clinical Images Using Deep Learning

M.H. Jafari[1], N. Karimi[1], E. Nasr-Esfahani[1], S. Samavi[1,2], S.M.R. Soroushmehr[3], K. Ward[4], K. Najarian[2,3]

[1]Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, 84156-83111 Iran
[2]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, 48109 U.S.A.
[3]Michigan Center for Integrative Research in Critical Care, University of Michigan, Ann Arbor, 48109 U.S.A.
[4]Department of Emergency Medicine, University of Michigan, Ann Arbor, 48109 U.S.A.

***Abstract*— Melanoma is the most aggressive form of skin cancer and is on rise. There exists a research trend for computerized analysis of suspicious skin lesions for malignancy using images captured by digital cameras. Analysis of these images is usually challenging due to existence of disturbing factors such as illumination variations and light reflections from skin surface. One important stage in diagnosis of melanoma is segmentation of lesion region from normal skin. In this paper, a method for accurate extraction of lesion region is proposed that is based on deep learning approaches. The input image, after being preprocessed to reduce noisy artifacts, is applied to a deep convolutional neural network (CNN). The CNN combines local and global contextual information and outputs a label for each pixel, producing a segmentation mask that shows the lesion region. This mask will be further refined by some post processing operations. The experimental results show that our proposed method can outperform the existing state-of-the-art algorithms in terms of segmentation accuracy.**

Keywords—Melanoma; medical image segmentation; skin cancer; convolutional neural network; deep learning.

I. INTRODUCTION

Melanoma, caused by abnormal reproduction of melanocyte cells, is the deadliest form of skin cancer. Melanocytes cells are responsible for producing melanin pigments that give color to skin [1]. Incidence of melanoma has risen rapidly over the past 30 years. It kills an estimated of 10000 people in the USA every year [2]. Melanoma, if detected in its early stages of growth, is highly curable [3]. There are various methods in dermatology for early diagnosis of melanoma, such as the criterion of ABCD (asymmetry, border irregularity, color patterns, and diameter) [4], and the seven-point checklist [5].

There is an ongoing active research for computer assisted analysis and diagnosis of melanoma. A review of existing research efforts is given in [6] and [7]. Recently there has been a rising trend for detection of melanoma in skin images captured by conventional user grade digital cameras (namely non-dermoscopic or clinical images). Dermoscopic images produced by dermoscope, a special purpose dermatology instrument, usually have uniform illumination and also have more contrast. On the other hand, the non-dermoscopic clinical images have the advantage of broad availability.

An important step in computerized analysis of melanoma is to locate the exact region of pigmented lesion in a skin image. This means that the image has to be segmented into two regions as lesion and normal skin. Performances of other steps that follow, such as feature extraction and classification, are directly depended on the accuracy of the segmentation step. Quality of segmentation can highly affect the assessment of metrics such as border irregularity and lesion asymmetry. There are many algorithms for automatic extraction of lesion region in skin images. These methods can be categorized in three groups [8] of thresholding, active contours, and region merging methods [9]. A review of existing methods is given in [8] and [10].

Standard digital images of skin are more challenging to consider due to containing artifacts such as uneven illumination or reflection and presence of noise. These effects mislead the common segmentation methods. Hence some existing segmentation algorithms are specifically designed for non-dermoscopic digital images [10-13]. Some methods consider color information in single channel [11], all channels [12] or efficient channel set found by principle component analysis [13]. Work of [10], one of the most recent researches in this area, has noticed the texture distinctiveness between lesion and normal skin. The applied segmentation methods usually have a preprocessing step, which deals with illumination effects and artifacts [14]. These methods also have some post processing steps for refinement of the segmentation mask.

Recently, deep learning methods have shown promising results in various pattern recognition and medical imaging applications. A review of literature and applications is given in [15] and [16]. Convolutional neural networks (CNN) are among widely used deep learning methods, which are inspired by human visual cortex. CNNs have been successfully adopted for numerous imaging applications. In [17], a CNN configuration is proposed for detection of salient regions in images where each super pixel is labeled based on its local and global context. Research efforts of [18-20] are recent examples of using deep neural networks for various medical image segmentation applications. In [18] a CNN is used for extraction of vessels in fundus images and [19] and [20] are applications of CNN for brain tumor segmentation.

In this paper, we propose a deep CNN architecture for extraction of lesion region from skin images. The input images are generated by standard cameras; hence, they should be preprocessed in order to handle noisy artifacts. For this purpose, we have proposed to enhance the image by applying a guided filter, reducing effects of disturbing elements on the accuracy of segmentation. This specific filtering procedure enhances the performance of the algorithm. Afterward a local patch is defined as a window around each pixel of the image, showing the local texture around the center pixel. In addition, a zoomed-out window, with the same center as the local patch, is used for revealing the global structure of the region. The two patches of local and global texture are fed to the proposed CNN. About 140,000 patches are obtained and used for the training of the CNN. The experimental results show that our proposed system can outperform other state-of-the-art methods in terms of lesion segmentation accuracy.

The rest of this paper is organized as follow. In section II the proposed method is discussed. The experimental results of qualitative and quantitative evaluation of our method are presented in section III. Finally section IV concludes the paper.

II. Proposed Method

In this section the proposed method for automatic segmentation of skin lesions from non-dermoscopic images is explained in details. Our method consists of four steps which are summarized in Fig. 1. In the rest of this section these steps are discussed.

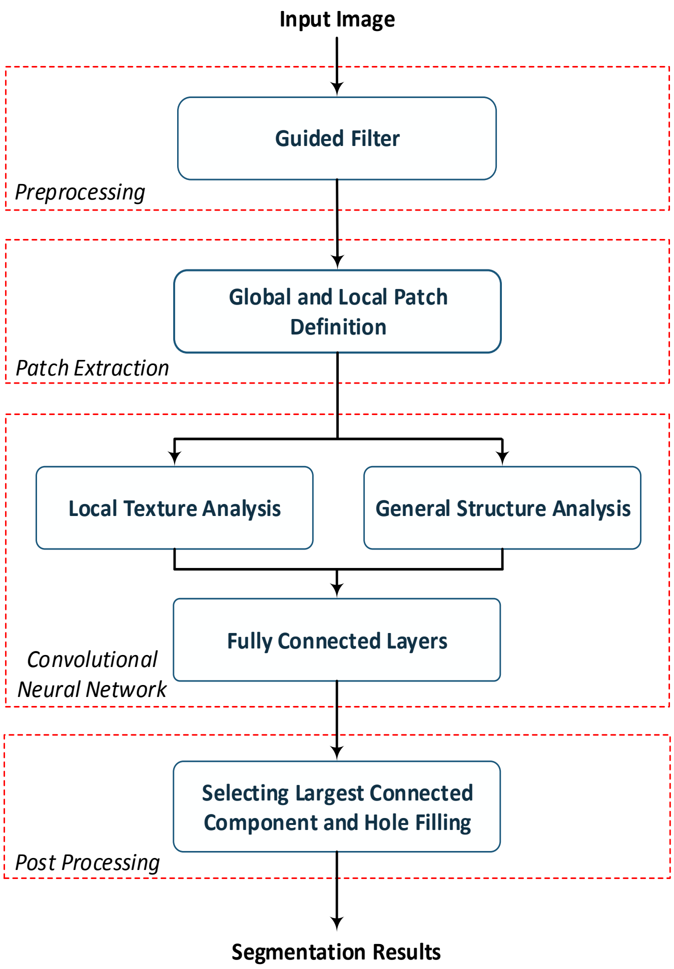

Fig. 1. Block diagram of four steps of the proposed method.

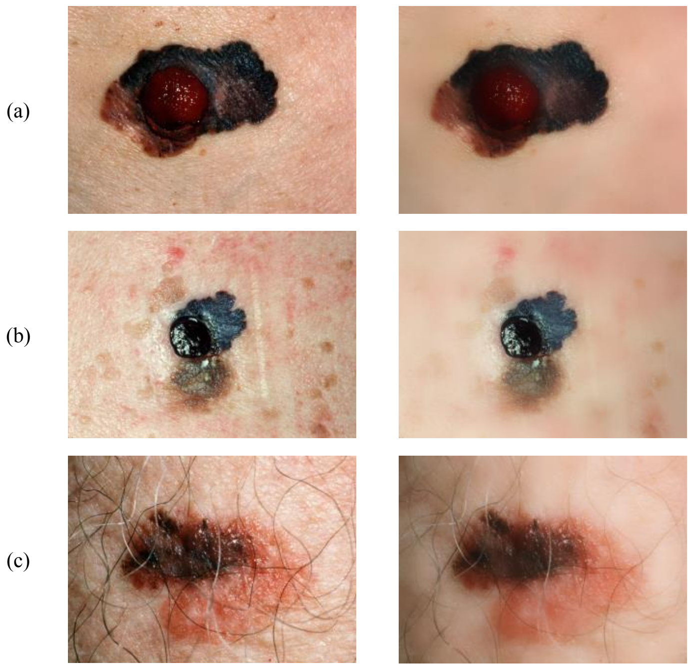

Fig. 2. Left column are input images. Right column are preprocessed images, resulted from applying guided filter.

A. Pre Processing

The input images of skin usually contain noisy artifacts such as hair, light reflection from skin surface and uneven illumination. These factors can have negative impact on segmentation performance and require handling. For this aim, a guided filter is applied on the input image as a preprocessing stage. Guided filter [21] is used as an edge preserving smoothing operator. It reduces the effect of noisy factors while causing minimal distortion on lesion's border. The input image itself is used as guidance image for the guided filter computation. In Fig. 2 the effect of applying guided filter, on some sample images from the dataset, is shown. The shown images have 1640×1043 pixels and the size of the guided filter neighborhood is set to 100. Noisy effects and irrelevant local textures, that could misdirect segmentation procedure, are smoothed out, while borders between lesion and normal skin regions are preserved by the guided filter operator. This preprocessed image is fed to a CNN with an architecture that is explained in the following.

B. Patch Extraction and CNN Architecture

To decide whether a pixel in the image belongs to lesion region or normal skin, as a patch is considered by placing a window around the pixel. These patches are fed to the used CNN and the output would label the pixel in the center of the patch. A patch with small window size would show local texture around the pixel. Local information can be beneficial in accurate detection of the lesion's border. Meanwhile the global context of the region, as where the pixel is located, is also important for decision about pixel's label. A lesion on a skin surface can be easily identified in a global view. In some applications, consideration of local and global contexts simultaneously improves the segmentation accuracy of CNN as shown in deep networks of [17] and [19].

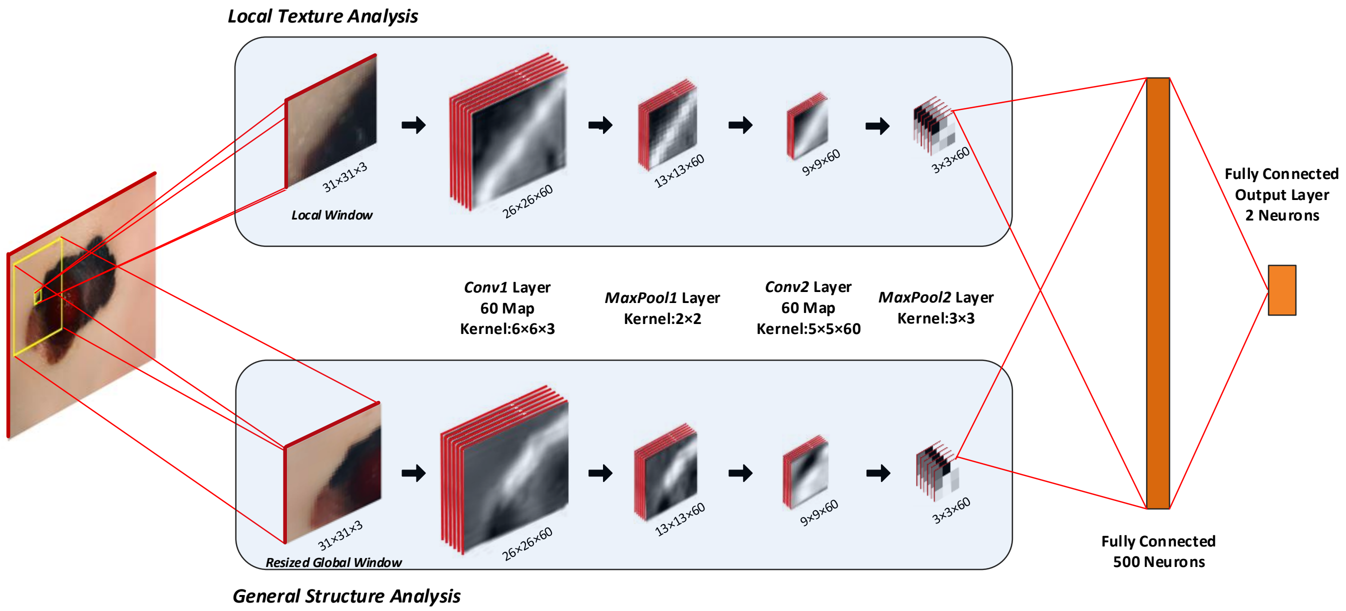

Fig. 3. Architecture of the proposed CNN.

The used CNN architecture is shown in Fig. 3. In our method, the image is resized to 600×400. Afterward, a local patch of size 31×31 is defined around each pixel of the image. Moreover, a window of size 201×201 is defined with the same center as the center of the local patch. This larger patch will also be resized to same size as the local patch, i.e. 31×31. In this global patch the general structure of the pixel's region is visible. In case a part of the patch falls outside of the image, the input image is padded. The padding is done by reflecting the image across 4 directions of left, right, up, down and along the directions of the four corners. By extraction of the local and global patches around each pixel, two patches are fed into two parallel CNNs with similar layers. At the end, fully connected concatenation layers are used where the results of analysis of the local and global textures are combined to form the final decision about the label of the central pixel in the patch.

As can be seen in Fig. 3, each wing of the CNN consists of four layers with the following order: *Conv1*, *MaxPool1*, *Conv2*, and *MaxPool2*. The first convolutional layer has a kernel size of $6 \times 6 \times 3$ and the kernel size in the second convolve layer is $5 \times 5 \times 60$. Each convolutional layer consists of 60 feature maps. Each feature map in a convolutional layer detects a single kind of feature across entire image.

The two convolutional layers are followed by max pooling layers with kernel size of 2×2 and 3×3 respectively. Usually in CNNs, convolve layers are followed by pooling layers. The pooling layers can facilitate the procedure of learning by reducing the number of variables that should be learned in the network. It will be done by ignoring the exact positional information of the extracted features. In the used CNN, the layers along two wings of the network, which process local and global patches, have same configurations.

Finally, there is a fully connected layer with 500 neurons, followed by the output $softmax$ layer with 2 neurons. It produces the probability of pixel's membership in lesion and normal skin regions.

C. Post Processing

By applying each patch of the input image to the explained CNN a segmentation map for the image will be obtained. In this map, each point is labeled by the class with higher probability of membership. The map has label 1 if the pixel is lesion or it is labeled as 0 if the pixel belongs to normal skin region. At the end the output mask is refined by selection of largest connected component, since it is assumed that there is a single lesion in each image. Then a hole-filling morphological operation is applied. The segmentation results of our method and quantitative evaluations are given in the following section.

III. Experimental Results

In this section, the performance of our method is evaluated on a dataset of skin lesion images from Dermquest database, which is publically available with segmentation ground-truth [22]. The used dataset consists of 126 digital images (66 melanoma, 60 non-melanoma). The proposed method is implemented in Matlab and Caffe [23], on a system with Intel Core i7-4790K processor, 32 GB of RAM, and NVIDIA GeForce GTX Titan X GPU card. The experiment is done as a leave-p-out cross-validation. The dataset is randomly split into four equal size groups. The experiment is done on one group that is left for test and three other groups are used as training,

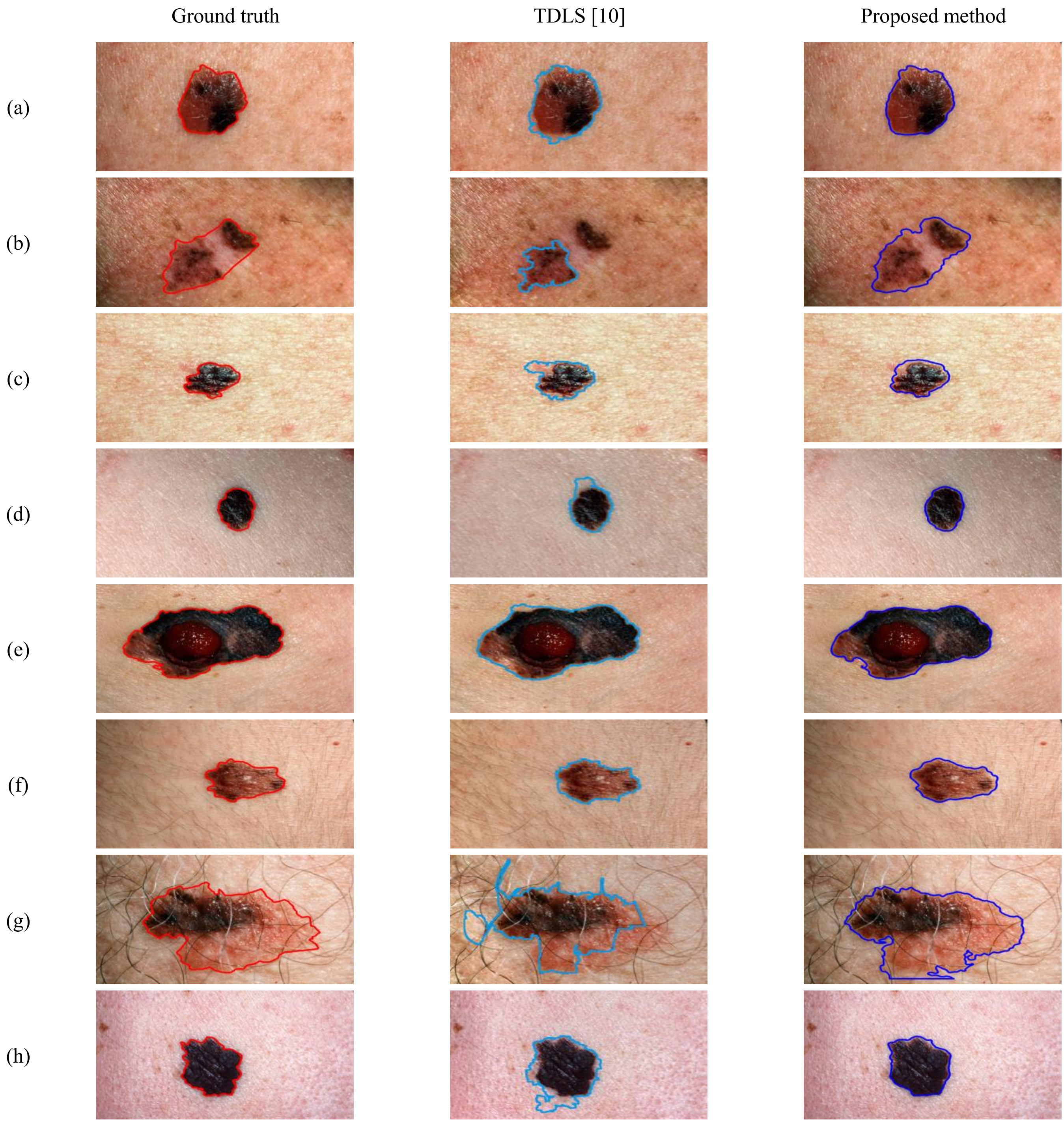

Fig. 4. Sample skin lesion segmentation results. Left column: input images superimposed by segmentation ground-truth (red line). Midle column: images of results of TDLS method [10]. Right column: masks resulted from our method.

i.e. the train and test ratio is 75% to 25%. The procedure is repeated four times for all different test groups.

The number of patches that are selected from each training image is 1500, where half of them are randomly chosen from lesion region and the other half are randomly selected from non-lesion parts. Hence, a total of about 140,000 randomly selected patches are extracted and used for training of the CNN in each run. The solver type is stochastic gradient descent (SGD). Also, $Xavier$ method is used for weight initialization and bias values are initialized to zero.

A. Qualitative Evaluation

In Fig. 4, ground-truth of some sample images are presented. Also, results of the proposed method are compared with TDLS algorithm [10], which is one of the state-of-the-art methods. Images of TDLS are obtained from [24]. Complex skin patterns that appear similar to lesions and some skin artifacts can mislead any segmentation algorithm. As can be seen in Fig. 4 (a), (c) and (h), the TDLS method has failed to accurately extract lesions' borders and it suffers from too many false positive pixels. In some images such as Fig. 4 (b) and (g), complicated skin texture and presence of artifacts such as hair have prevented TDLS from correctly detecting the lesion's region. However our method has a high qualitative performance even in these challenging situations.

B. Quantitative Evaluation

For quantitative evaluation of our method, three commonly used metrics for classification problems are used. These metrics are sensitivity, specificity and accuracy. The segmentation problem can be stated as a classification problem, where each pixel is classified into one of the two classes of lesion (positive) and normal skin (negative). As a result, by comparing the output mask with the ground-truth, the classification metrics can be calculated as:

$$sensitivity = \frac{TP}{TP+FN},$$

$$specificity = \frac{TN}{TN+FP},$$

$$accuracy = \frac{TP+TN}{TP+FN+TN+FP},$$

where TP, TN, FP and FN denote the number of true positive, true negative, false positive and false negative classified pixels respectively.

The proposed method is compared with five other methods based on reported results of the mentioned metrics on the same dataset. The results are shown in Table 1. As can be seen, the proposed method can reach the best performance in segmentation sensitivity and accuracy in comparison with other state-of-the-art methods. Also in terms of specificity we achieved a relatively high score.

TABLE I. QUANTITATIVE COMPARISION OF LESION SEGMENTATION RESULTS, BEST RESULTS ARE BOLDED.

Segmentation Algorithm	Segmentation Performance		
	Sensitivity	***Specificity***	***Accuracy***
L-SRM [9]	89.4	92.7	92.3
Otsu-R [11]	87.3	85.4	84.9
Otsu-RGB [12]	93.6	80.3	80.2
Otsu-PCA [13]	79.6	**99.6**	98.1
TDLS [10]	91.2	99.0	98.3
Proposed Method	**95.0**	98.9	**98.5**

IV. CONCLUSION

Proper segmentation of skin cancer images, for accurate identification of lesion region, is of great importance. Segmentation accuracy can highly affect the next steps of the diagnosis. In this paper a method based on deep convolutional neural networks was proposed for extraction of lesion region in digital clinical images. All input images are initially preprocessed by applying an edge preserving smoothing guided filter for reducing noisy artifacts. Then each pixel of the preprocessed image, as a center of a patch, is fed to a CNN. Two patches, with local and global natures, are formed around each pixel and are fed into a CNN. The output of the CNN is a label for the center pixel of the patch. The proposed preprocessing filter and the proposed CNN structure were highly suitable for this critical segmentation procedure. Experimental results showed that the proposed method can reach a very high accuracy of 98.5% and sensitivity of 95.0% that outperforms other state-of-the-art methods.

REFERENCES

[1] American Cancer Society, "Cancer Facts & Figures 2016," *American Cancer Society*, Atlanta, GA, USA, 2016.

[2] Skincancer.org, "Melanoma - SkinCancer.org," 2016. [Online]. Available: http://www.skincancer.org/skin-cancer-information/.

[3] A. F. Jerant, J. T. Johnson, C. Sheridan and T. J. Caffrey, "Early detection and treatment of skin cancer," *American family physician*, no. 2, pp. 357-386, 2000.

[4] F. Nachbar, W. Stolz, T. Merkle, A. B Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. B.-Falco, and G. Plewig, "The abcd rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551–559, 1994.

[5] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comp. of the abcd rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Archives of Dermatology*, vol. 134, no. 12, pp. 1563-1570, 1998.

[6] I. Maglogiannis and C. Doukas, "Overview of Advanced Computer Vision Systems for Skin Lesions Characterization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 721-733, 2009.

[7] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artificial intel. in medicine*, pp. 69-90, 2012.

[8] M. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 33, no. 2, pp. 148–153, 2009.

[9] M. Emre Celebi, H. A. Kingravi, H. Iyatomi, Y. Alp Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz, and S. W. Menzies, "Border detection in dermoscopy images using statistical region merging," *Skin Research and Technology*, vol. 14, pp. 347-353, 2008.

[10] J. Glaister, A. Wong, and D. A. Clausi, "Segmentation of Skin Lesions From Digital Images Using Joint Statistical Texture Distinctiveness,"

IEEE Transactions on Biomedical Engineering, vol. 61, pp. 1220-1230, 2014.

[11] P. Cavalcanti, Y. Yari, and J. Scharcanski, "Pigmented skin lesion segmentation on macroscopic images," in *Proc. 25th International Conference on Image Vision Computing*, pp. 1–7, 2010.

[12] P. G. Cavalcanti, J. Scharcanski, and C. Lopes, "Shading Attenuation in Human Skin Color Images," in *Advances in Visual Computing: 6th International Symposium*, pp. 190-198, 2010.

[13] P. G. Cavalcanti and J. Scharcanski, "Automated prescreening of pigmented skin lesions using standard cameras," *Computerized Medical Imaging and Graphics*, Elsevier, no. 6, pp. 481-491, 2011.

[14] J. Glaister, R. Amelard, A. Wong, and D. A. Clausi, "MSIM: Multistage Illumination Modeling of Dermatological Photographs for Illumination-Corrected Skin Lesion Analysis," *IEEE Transactions on Biomedical Engineering,* vol. 60, pp. 1873-1883, 2013.

[15] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85-117, 2015.

[16] L. Deng and D. Yu, "Deep Learning: Methods and Applications," *Found. Trends Signal Process.*, vol. 7, pp. 197-387, 2014.

[17] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1265-1274, 2015.

[18] M. Melinščak, P. Prentašić, and S. Lončarić, "Retinal Vessel Segmentation Using Deep Neural Networks," in *10th International Conference on Computer Vision Theory and Applications*, 2015.

[19] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain Tumor Segmentation with Deep Neural Networks," *arXiv preprint arXiv:1505.03540*, 2015.

[20] S. Pereira, A. Pinto, V. Alves and C. Silva, "Deep Convolutional Neural Networks for the Segmentation of Gliomas in Multi-sequence MRI," *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 131-143, 2016.

[21] K. He, J. Sun, and X. Tang, "Guided Image Filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1397-1409, 2013.

[22] [Online].Available:http://vip.uwaterloo.ca/demos/skin-cancer-detection/.

[23] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding," in *Proc. of the 22nd ACM international conference on Multimedia*, pp. 675-678, 2014.

[24] J. Glaister, "Automatic segmentation of skin lesions from dermatological photographs," *Department of Systems Engineering*, University of Waterloo, Waterloo, ON, Canada, 2013.