SEGMENTATION OF DERMOSCOPY IMAGES BASED ON FULLY CONVOLUTIONAL NEURAL NETWORK

Zilin Deng [1,2] *Haidi Fan, Fengying Xie*[*2] *Yong Cui*[3] *Jie Liu*[4]

[1] School of Biological Science and Medical Engineering, Beihang University, Beijing 100083, China
[2] Image Processing Center, Beihang University, Beijing 100083, China
[3] Department of Dermatology and Venereology, China-Japanese Friendship Hospital, Beijing 100029, China
[4] Department of Dermatology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100032, China

ABSTRACT

Lesion segmentation is one of the crucial steps for computerized dermoscopy image analysis. To accurately extract lesion borders from dermoscopy images, a novel segmentation method based on fully convolutional neural network is proposed in this paper. The designed network contains a low-level trunk followed by two brunches (global brunch and local brunch). The low-level trunk is fine-tuned from VGG16 net. Two brunches with different receptive fields extract global and local features respectively. After the combination of the global and local features, the final segmentation results are obtained through pixel-wise softmax classification. Experiments are conducted on the challenge dataset ISBI 2016. The results demonstrate that our designed network is more adaptive to dermoscopy images, which obtain more accurate lesion borders with good robust than other state-of-the-art methods.

Index Terms— Lesion segmentation, Dermoscopy, Fully convolutional neural network.

1. INTRODUCTION

Melanoma is believed as the most deadly form of skin cancer [1]. Dermoscopy is a non-invasive diagnosis technique for malignant melanoma (MM). It enables the generation of images with constant illumination, different texture patterns, and characteristics that are sometimes difficult to detect by simple naked-eye. To assist dermatologist in making correct diagnosis for MM on early stage, many researchers concentrate on the computerized dermoscopy image analysis system for proper interpretation of dermoscopy images. A computerized dermoscopy image analysis system includes five steps: image acquisition, pre-processing, lesion segmentation, feature extraction, and classification. Lesion segmentation is a crucial step in analysis system.

Lesion segmentation is quite challenging due to the complexity of dermoscopy images: 1) low contrast between lesion and healthy skin; 2) irregular and fuzzy lesion borders; 3) hairs; 4) variegated color; 5) fragmentation made by depigmentation; 6) incomplete lesion. It is reported that 28% of the computer vision articles dating from 1984 to 2012 on single lesion analysis is concentrated on border detection [2].

Numbers of methods have been developed for this challenging problem. Most of them can be broadly classified to three categories: thresholding [3][4], active contour [5][6], and region merging methods [7][8]. Thresholding methods can achieve good results when there is good contrast between lesion and skin, but fail to detect lesion accurately if the contrast is not explicit. Active contour methods perform poorly when the boundaries are not defined clearly and hairs overlap the lesion area. Region merging methods have difficulties when the lesion and the skin region are textured or have variegated colors. In addition, conventional segmentation methods usually utilize specifically-designed pre-processing and post-processing steps to deal with illumination effects, hairs and artifacts [9]. However, these specifically-designed steps may lower the robustness of methods in clinic diagnosis, considering the complexity of dermoscopy images.

Recently, convolutional neural networks (CNNs) have shown its promising results and powerful generalization ability in many medical image analysis tasks, such as segmentation [10][11] and classification [12][13]. An overview and future promise about the application of CNNs in medical image analysis tasks is given in [14]. CNNs have been successfully applied in dermoscopy image analysis approaches. Codella et al. [15] utilized CNNs to extract high-level features for melanoma classification. In [16], a fully convolutional residual network (FCRN) is constructed for accurate skin lesion segmentation.

The main contributions of our work can be summarized as follows:

1) We propose a novel dermoscopy segmentation method based on fully convolutional neural network (FCN). Our network have two branches with different size of receptive field, extracting local and global features respectively. This architecture gives consideration to both segmentation accuracy and fine edges. Compared with most dermoscopy segmen-

*Email: xfy_73@buaa.edu.cn. This work was supported by the National Natural Science Foundation of China (Grant 61471016).

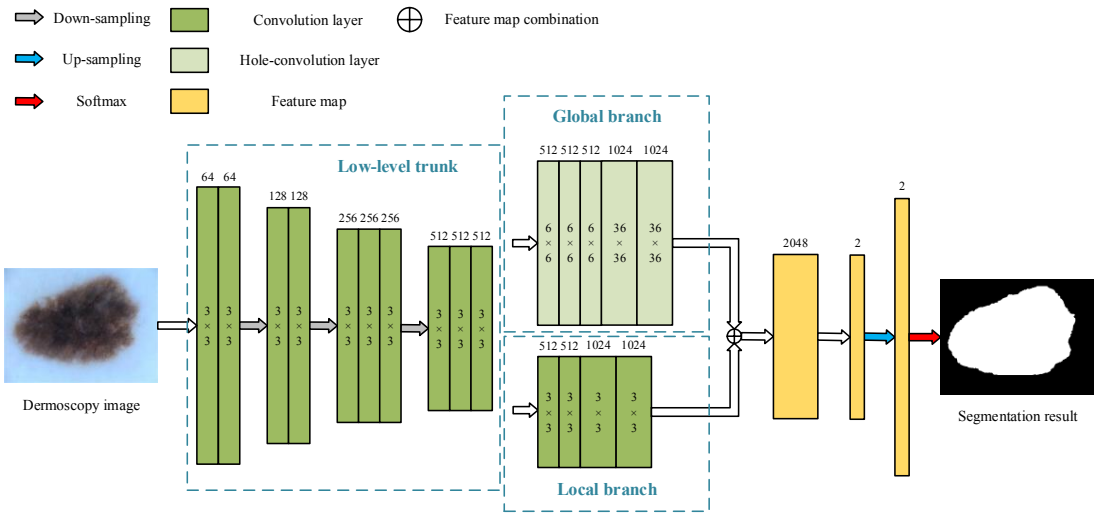

Fig. 1. Architecture diagram of our two-branch FCN. The numbers in blocks and above blocks represent the size of convolutional filters and the number of feature maps, respectively. The down-sampling is performed by 3×3 max-pooling (stride is 2) and the up-sampling is performed by bilinear interpolation.

tation methods, our framework is end-to-end, which means that the framework is entirely trainable and there are no hand-designed and fixed module that cannot learn from data. The end-to-end design enhances the robustness of our method.

2) We validate the applicability of transfer learning on dermoscopy images. Various state-of-the-art fully convolutional neural networks are transferred and compared in experiment stage.

The remainder of this paper is organized as follows. In section 2, we describe our proposed approach in detail. Experimental results are presented and discussed in section 3. Section 4 gives the conclusion of our work.

2. METHOD

CNNs was proposed by Lecun in 1989 [17] and its incredible power have been proven in many recognition tasks. In 2015, Long et al. [18] developed the new CNNs architecture called FCN, by transforming the fully connected layer of CNNs into convolutional layer. This transformation enables FCN to complete segmentation tasks on arbitrary size of images. In this paper, a two-branch FCN architecture is designed for the segmentation of dermoscopy images, as shown in Fig. 1.

2.1. Low-level trunk

As mentioned above, lesion segmentation is a challenging task due to the complexity of dermoscopy images. To deal with various situations, the designed network should be deep enough to extract more expressive features. However, considering the restricted size of the public dermoscopy dataset, a randomly initialized deep network has long learning time and difficulty of convergence. Therefore, we employ transfer learning method to reduce learning time and improve network robustness. The most common transfer learning method used in deep learning is fine-tuning.

The low-level layers of our network are fine-tuned from the top ten layers of VGG16 net [19], a classic deep learning network proposed for classifying natural scene images. The powerful expression of features extracted from VGG16 net has been verified in many FCN architectures [18][20].

2.2. Global and local branches

When a FCN goes deep, a trade-off problem between classification accuracy and localization accuracy become a crucial limitation on the performance of the network [20]. Deeper models with multiple down-sampling layers (such as max-pooling layer) and large receptive field have proven successful in classification tasks. However, their high built-in invariance to local image transformations weaken their abilities in inferring position from output. To balance the classification ability and localization ability for our network, we design a two-branch architecture including global branch and local branch, exacting global and local features respectively.

The global branch is fine-tuned from the 11th to 15th layers of VGG16 net. Two max-pooling layers are removed for avoiding down-sampling. Furthermore, to obtain global information for network, we enlarge the receptive field of those layers by hole algorithm [20] ($2\times$ in the first three layers and $12\times$ in the last two layers). Hole algorithm modifies the convolutional filters in the layers by filling zero between each

original filter parameter [21], through which we can implement fine-tuning even though the length of convolutional filters have been changed. In this way, the global branch offers global features for our network without the loss of localization ability occurred by down-sampling.

The local branch consists of four convolution layers. Because of its simple architecture, the filter parameter of the local branch is randomly initialized (not fine-tuned). Comparing to the global branch, the local branch has smaller receptive field and pays more attention on the small area lesion. We combine global and local feature maps together for prediction and up-sample the prediction probability to the size of the original image by bilinear interpolation. The final segmentation results are generated by pixel-wise softmax classification.

Figure 2 presents two examples of the combination of global and local branches. In the first row of Fig.2, a small area lesion is missed by the global branch because of the large receptive field, while the local branch detects it but with under-segmentation. When combining the two branches together, we obtain more accurate lesion border. In the second row of Fig.2, a dark region in the artifact is wrongly detected as a lesion by the local branch because of its weak classification ability caused by the small receptive field, whereas the global branch can accurately determine whether a big area region is a lesion. Therefore, through the combination of global and local branches, our designed network can detect small area lesion and at the same time, avoid from over-segmentation.

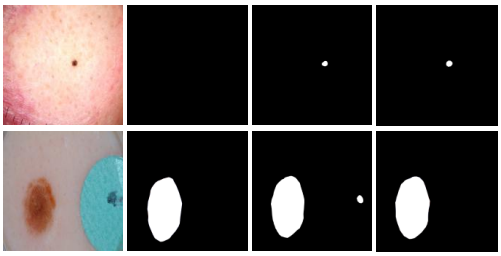

Fig. 2. Segmentation results of two dermoscopy images. (a) original image; (b) result by global branch; (c) result by local branch; (d) final result.

3. EXPERIMENT

We evaluated the performance of our method on a public challenge dataset of ISBI 2016 [22], which was based on International Skin Imaging Collaboration (ISIC) Archive. It is the largest publicly available dataset on dermoscopy images currently, including 900 images as training data and 379 images as testing data, and the ground truth is also provided by the organizer.

Five metrics were used to quantify segmentation performance, including Jaccard index (JA), accuracy (AC), Dice coefficient (DI), sensitivity (SE) and specificity (SP). They are defined as:

$$JA = \frac{TP}{TP + FP + FN}, \tag{1}$$

$$AC = \frac{TP + TN}{TP + FP + FN + TN}, \tag{2}$$

$$DI = \frac{2TP}{2TP + FP + FN}, \tag{3}$$

$$SE = \frac{TP}{TP + FN}, \tag{4}$$

$$SP = \frac{TN}{TN + FP}, \tag{5}$$

where TP, TN, FP and FN denote the pixel number of true positive, true negative, false positive and false negative, respectively. "Positive" and "negative" are defined by the ground truth.

3.1. Comparison with other network architectures

To verify the superiority of our two-branch network and the effectiveness of transfer learning, we compared the segmentation performance of the proposed network with two state-of-the-art FCN networks (FCN-32s [18] and deeplab [20]) and their derivative structures, including large field of view (LFOV), multiscale (MSc) and conditional random fields (CRF).

All networks were trained by stochastic gradient descent (SGD) with momentum 0.9 and weight decay 0.005. For the proposed two-branch network and two single-branch networks, we used a minibatch size of 4 images and set learning rates as 0.001 initially and reduced it by a factor of 100 every 4000 iterations. For other compared networks, learning rates were set according to their original literature. Figure 3 presents some example results of different networks, where red line represents the ground truth and blue line indicates the segmentation result. It can be seen that, for challenging images (with hair and artifact), Deeplab + LFOV + CRF, Deeplab + LFOV + MSc + CRF and our method are able to achieve more accurate results. CRF [23] integrates the RGB value and position of each pixel with the prediction probability to avoid from over-segmentation, as shown in the second row of Fig. 3 (d)-(g). However, hair has similar color with the lesion and often overlaps on the lesion, which causes wrong segmentation result. Therefore, compared with other FCN architectures, our designed network is insensitivity to complex image capture condition and can obtain more accurate lesion borders.

Table 1 is the statistic results of different networks on 379 test images. All the FCN networks obtained good results, which validated the applicability of transfer learning

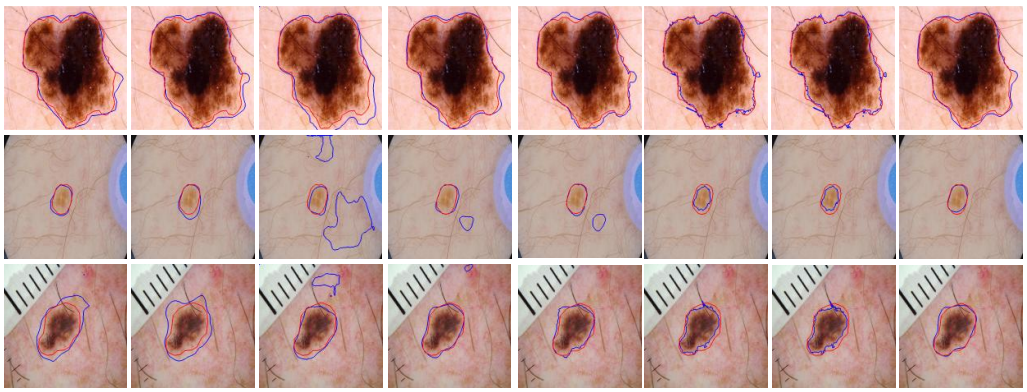

Fig. 3. Example results of different networks. (a) FCN-32s; (b) FCN-8s; (c) Deeplab; (d) Deeplab + LFOV; (e) Deeplab + LFOV + MSc; (f) Deeplab + LFOV + CRF; (g) Deeplab + LFOV + MSc + CRF; (h) Our method.

Table 1. Segmentation result statistics of different network architectures.

Method	JA (%)	AC (%)	DI (%)	SE (%)	SP (%)
Fcn-32s	82.1	94.4	89.1	91.5	95.0
Fcn-8s	82.5	94.6	89.3	92.1	94.8
Deeplab	77.6	93.2	86.1	92.9	91.9
Deeplab+LFOV	83.5	95.0	90.1	93.2	95.1
Deeplab+LFOV+MSc	83.5	95.1	89.8	93.0	95.2
Deeplab+LFOV+CRF	83.2	95.1	89.7	87.9	**96.6**
Deeplab+LFOV+MSc+CRF	83.1	95.0	89.5	87.9	96.5
Our method	**84.1**	**95.3**	**90.7**	**93.8**	95.2

in dermoscopy images. For FCN-32s and FOV-8s, they get similar segmentation accuracy. For Deeplab, LFOV can obviously improve the segmentation performance (see the fifth row in Tab.1). When using MSc, the segmentation accuracy almost does not change. CRF method not only can not improve the segmentation accuracy, but also cause a decrease in SE, Therefore, with the best AC, DI, JA and SE values, our designed network architectures outperforms other compared network architectures.

3.2. Comparison with other segmentation methods

Several research groups participated in the segmentation challenge on ISBI dataset in 2016. Both traditional methods and deep learning methods are utilized by those groups. We compared our approach with the best six methods in the challenge, which represent the state-of-the-art segmentation methods for dermoscopy images. Table 2 shows the comparison results, where the statistics values of comparison methods are directly from [24]. All the methods were ranked according to JA metric. Our method followed tightly to the first one EXB [25] and outperformed other five compared methods. EXB utilized some specifically-designed pre- and post- processing steps for refining segmentation results. However, as an end-to-end approach, our method does not use any additional process. Therefore, our method has greater potential in complex clinic diagnosis and achieves better segmentation results than the six comparison methods.

Table 2. Segmentation result statistics of different segmentation methods.

Method	JA (%)	AC (%)	DI (%)	SE (%)	SP (%)
EXB	**84.3**	**95.3**	**91.0**	91.0	96.5
Our method	84.1	**95.3**	90.7	**93.8**	95.2
CUMED	82.9	94.9	89.7	91.1	95.7
Mahmudur	82.2	95.2	89.5	88.0	96.9
SFU-mial	81.1	94.4	88.5	91.5	95.5
TMUteam	81.0	94.6	88.8	83.2	**98.7**
UiT-Seg	80.6	93.9	88.1	86.3	97.4

4. CONCLUSION

A novel segmentation method based on FCN is proposed for dermoscopy images in this paper. The designed FCN network consists of a low-level trunk and two brunches (global brunch and local brunch). To reduce learning time and improve network robustness, the low-level trunk is first fine-tuned from VGG16 net. In the two branches, the global brunch, whose receptive field is enlarged using hole algorithm, is employed to extract global feature maps for accurate pixel classification without down-sampling, and the local brunch with smaller receptive field is used to extract local feature maps for detecting small area lesion. Combining global and local feature maps together, we obtain the final segmentation results through pixel-wise softmax classification. Experimental results show that the designed FCN network is more robust for dermoscopy images with complex conditions than other FCN architectures, and the proposed segmentation framework achieves better segmentation results than the best six methods participated in ISBI 2016 challenge.

5. REFERENCES

[1] A. F. Jerant, J. T. Johnson *et al.*, "Early detection and treatment of skin cancer," American family physician, vol. 62, no. 2, pp. 357-386, 2000.

[2] K. Korotkov and R Garcia, "Computerized Analysis of Pigmented Skin Lesions: A Review," Artificial Intelligence in Medicine, vol. 56, no. 2, pp. 69-90, 2012.

[3] P. G. Cavalcanti and J. Scharcanski, "Automated prescreening of pigmented skin lesions using standard cameras," Computerized Medical Imaging and Graphics, vol. 35, no. 6, pp. 481-491, 2011.

[4] M. E. Celebi, Q. Wen *et al.*, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," Skin Research and Technology, vol. 19, no. 1, pp. e252-e258, 2013.

[5] Q. Abbas, I. Fondn and M. Rashid, "Unsupervised skin lesions border detection via two-dimensional image analysis," Computer methods and programs in biomedicine, vol. 104, no. 3, pp. e1-e15, 2011.

[6] R. Kasmi, K. Mokrani *et al.*, "Biologically inspired skin lesion segmentation using a geodesic active contour technique," Skin Research and Technology, vol. 22, pp. 208-222 2015.

[7] F. Xie and A. C. Bovik, "Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm," Pattern Recognition, vol. 46, no. 3, pp. 1012-1019, 2013.

[8] M. E. Celebi, H. A. Kingravi *et al.*, "Border detection in dermoscopy images using statistical region merging," Skin Research and Technology, vol. 14, no. 3, pp. 347-353, 2008.

[9] J. Glaister, R. Amelard *et al.*, "MSIM: Multistage Illumination Modeling of Dermatological Photographs for Illumination-Corrected Skin Lesion Analysis," IEEE Transactions on Biomedical Engineering, vol. 60, pp. 1873-1883, 2013.

[10] M. Havaei, A. Davy *et al.*, "Brain tumor segmentation with deep neural networks," Medical image analysis, vol.35, pp. 18-31, 2017.

[11] A. Prasoon, K. Petersen *et al.*, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in International conference on medical image computing and computer-assisted intervention, pp. 246-253, 2013.

[12] H. I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 583-590, 2013.

[13] S. Demyanov, R. Chakravorty *et al.*, "Classification of dermoscopy patterns using deep convolutional neural networks," in International Symposium on Biomedical Imaging, pp. 364-368, 2016.

[14] H. Greenspan, B. van Ginneken and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1153-1159, 2016.

[15] N. Codella, J. Cai *et al.*, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," International Workshop on Machine Learning in Medical Imaging, pp. 118-126, 2015.

[16] L. Yu, H. Chen *et al.*, "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks," IEEE Transactions on Medical Imaging, DOI: 10.1109/TMI.2016.2642839, 2016.

[17] Y. LeCun, B. Boser *et al.*, "Backpropagation applied to handwritten zip code recognition," Neural computation, vol. 1, no. 4, pp. 541-551, 1989.

[18] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440, 2015.

[19] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[20] L. C. Chen, G. Papandreou *et al.*, "Semantic image segmentation with deep convolutional nets and fully connected crfs," arXiv preprint arXiv:1412.7062, 2014.

[21] S. Mallat, "A wavelet tour of signal processing," Academic press, 1999.

[22] D. Gutman, N. C. F. Codella *et al.*, "Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)," arXiv preprint arXiv:1605.01397, 2016.

[23] V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," Adv. Neural Inf. Process. Syst, vol. 2, no. 3, pp. 109-117, 2011.

[24] `https://challenge.kitware.com/#phase/566744dccad3a56fac786787`

[25] K. Sirinukunwattana, J. P. Pluim *et al.*, "Gland segmentation in colon histology images: The glas challenge contest," arXiv preprint arXiv:1603.00275, 2016.