

# Skin Lesion Segmentation Based on Modification of SegNet Neural Networks

Quoc Cuong Ninh, Thi-Thao Tran, Tien Thanh Tran, Thi Anh Xuan Tran, Van-Truong Pham\*

*School of Electrical Engineering*

*Hanoi University of Science and Technology*

Hanoi, Vietnam

Email: truong.phamvan@hust.edu.vn

**Abstract**— Skin lesion segmentation plays an important role in automatic skin cancer diagnosis. One of the most dangerous types of this cancer is melanoma which requires an early and accurate detection. However, automatic melanoma segmentation on dermoscopic images is a challenging task since images are corrupted by noise like hairs, air bubbles, blood vessel... and with fuzzy boundaries. This paper presents a framework based on deep fully convolutional neural networks to automatically segment skin lesions in dermoscopic images. Particularly, the paper proposes a fully convolutional network (FCN) framework that is based on modification of the SegNet architecture. In particular, we propose to reduce the downsampling and upsampling layers in the original SegNet model, that reduces total learned parameters compared to the original SegNet model. The proposed approach is applied to segment images from ISIC 2017 dataset. Experimental results show the desired performances of the proposed approach in terms of metrics of Dice coefficient and Jaccard indexes.

**Keywords**— Skin lesion segmentation, Deep learning, Image segmentation, Skin cancer, Deep neural networks

## I. INTRODUCTION

Skin cancer is the most frequent type of cancer [1]. There are 5.4 million new cases confirmed in the United States every year [2]. Early detection and treatment might help to prevent the disease development [3]. Automated skin lesion segmentation from dermoscopic images is a crucial step in automated diagnosis and computerized analysis of dermoscopic images [4]. However, this is a challenging task due to the large variety in size, shape, and color along with different types of skin and texture [1, 5]. In addition, some lesions have irregular and fuzzy borders, and in some cases the contrast between lesion and the surrounding skin is pretty low. Moreover, artifacts and intrinsic cutaneous features, such as hairs, frames, blood vessels and air bubbles might make the automatic segmentation more challenging [1].

There exist various methods for skin lesions segmentation in the literature such as region merging [6], thresholding [7], and active contour based method [8]. In particularly, to tackle skin lesion segmentation problem, Celebi et al. [7] proposed a method for skin lesion segmentation by ensembling four different thresholding methods. Peruch et al. [5] proposed a segmentation method, in which the RGB images is firstly projected onto the first principal component of the color histogram, then a thresholding technique is used to cluster the colors into lesional and non-lesional skin. Zhou et al. [8] integrated a mass density function into the optimization objective functional to improve the performance of the classic GVF snake. Sadri et al. [9] introduced a fixed-grid wavelet networks in which orthogonal least squares was performed to

calculate the network weights and to optimize the network structure in a supervised way. Xie et al. [10] proposed a supervised method for skin lesion segmentation, where a self-generating neural network is combined with the genetic algorithms. In addition to the aforementioned unsupervised approaches, the supervised methods need to employ manually extracted features. Thus, leading to time consuming and requiring specialized domain knowledge of the experts. Moreover, in this approach, one needs to further pre-processing steps, such as contrast adjustment and enhancement, light and magnification, lesion localization to facilitate the lesion segmentation tasks [11].

On the other hand, recently, deep learning approaches for semantic segmentation have been developed and successfully demonstrated promising performances. A number of novel architectures have been recommended for the segmentation of medical images as well as for skin lesions. Fully convolutional network-based methods become popular and proved effectiveness in medical image segmentation. For example, U-net, an encoder-decoder architecture with skip-connection modified by Venkatesh et al. [12] and Ibtehaz et al. [13] has achieved significant results in skin lesion segmentation. Long et al. [14] designed a Fully Convolutional Networks (FCN) and Phi Vu Tran [15] has successfully applied FCN for cardiac segmentation in short-axis MRI. A combination of YOLO and GrabCut algorithm by Ünver and Ayan [16] has offered an alternative real-time skin lesion segmentation method.

In this study, we propose a framework based on deep fully convolutional neural networks to automatically segment skin lesions in dermoscopic images. The proposed FCN architecture is a modified version of SegNet neural network—the model has shown impressive results in many medical image segmentation tasks in recent years. The architecture consists of two main paths: encoder path, which is account for downsampling an image with a set of convolution and max-pooling layers, and the decoder path, which upsamples the feature maps to the same size of the initial image and preform a pixel-wise segmentation mask. We reduce the downsampling and upsampling layers in the original SegNet model, thus the total learned parameters of the model reduced while compared to the original model. Therefore, our model will be able to train from scratch. Experimental results on the ISIC 2017 dataset demonstrate that our model can obtained impressive result compared to other neural network-based methods. The results have been validated with two metrics: Dice and Jaccard coefficients. The average Dice score obtained by the proposed method for the database is of 0.853, and Jaccard similarity is of 0.771.

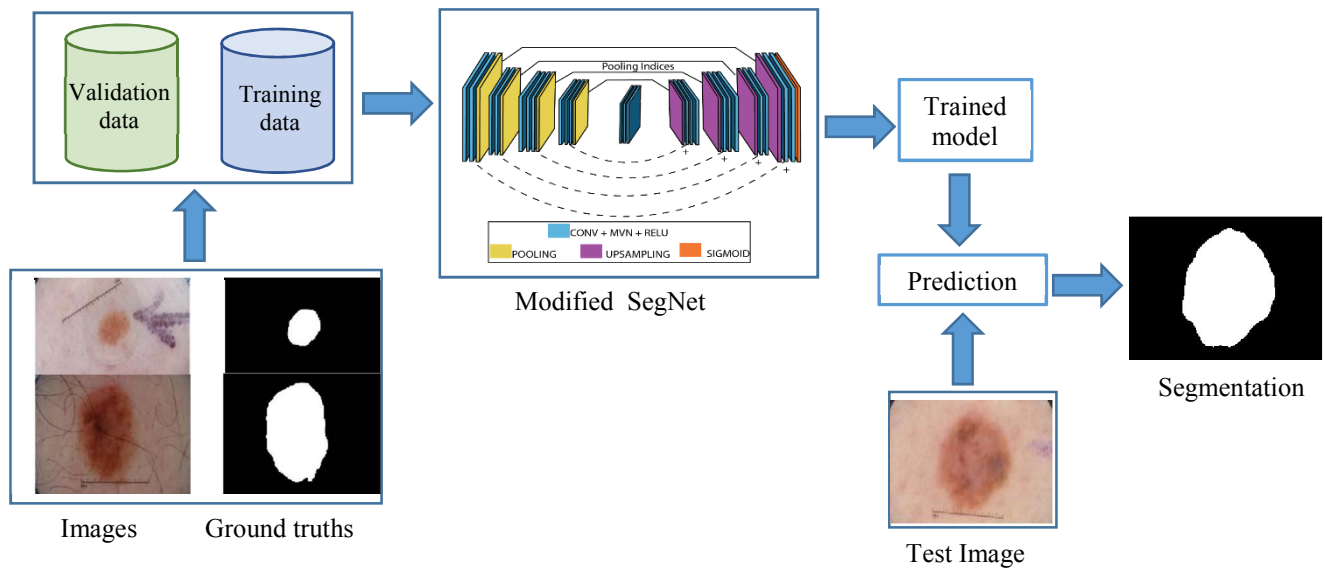


Fig.1 General architecture of the SegNet for skin lesion segmentation

## II. MATERIALS AND METHODS

### A. Network architecture

The proposed model is presented in Fig.1. The model is based on modified version of the SegNet architects. In particular, we propose to reduce the downsampling and upsampling layers in the original SegNet model for reducing the total learned parameters of the original SegNet model. The model therefore is not only computationally inexpensive, but also achieved outstanding results on the dermoscopic dataset.

The model is categorized into encoder and decoder parts. The encoder network of the model comprise 10 convolutional layers and the decoder network has 10 corresponding layers. Each encoder layers have small receptive field  $3 \times 3$ , one-pixel stride, various max pooling to preserve the spatial dimensions of the feature maps and followed by Rectified Linear Unit (ReLU) activation function and mean-variance normalization (MVN) operation. After some of convolution layers, a max-pooling which has  $2 \times 2$  pixel window, with stride 2 is performed. There are total 4 pooling layers all over the encoder network making the spatial dimensions of sample divided in two. However, the increasingly loss of spatial resolution of the feature maps leads to negative effect for segmentation when boundary detail is vanishing. So we can apply a simple technique and save the max-pooling indices, i.e., the locations of the maximum feature value in each pooling window to use in up-sampling layers in decoder network. By this way, we could store boundary information in the encoder feature maps in a negligible amount of memory.

In the decoder network, the model uses the max-pooling indices to up-samples the input feature maps. The output result of each up-sampling layer is added with the corresponding encoder feature map. Before this adding step, two feature maps are handled by  $1 \times 1$  convolutions to match dimensions. The result of adding steps are then convoluted by  $3 \times 3$  convolutional layers followed by MVN steps like the encoder network. The output of the decoder network is a feature map which has the same resolution and number of channels as the

input image. This output is fed to a trainable sigmoid activation function to classify each pixel independently.

### B. Training protocol

The optimal algorithm used is 'Stochastic Gradient Descent with Momentum' with the weight of moment equal to 0.9 to find the optimal point of the cross-entropy loss function-comparing the deviation of the predicted image with the label (ground truth) similarity. The weights of the network are initialized according to the method 'Xavier initialization'. Combining the 'MVN' operation after each ReLU with 'Xavier' initialization help to avoid 'vanishing/exploding gradient'-when the weights are too small or too large. The values at the nodes in deeper layer will be approximately to zero or infinity, so the weights will not be updated or the optimization cannot converge.

In order to avoid overfitting, we will use 'dropout' with a ratio of 0.5 and 'L2 regularization' with a weight of 0.0005. In addition, the 'learning rate decay' technique is also used by

the formula:  $\text{base\_lr} \times \left(1 - \frac{\text{iter}}{\text{max\_iter}}\right)^{\text{power}}$ , where  $\text{base\_lr} =$

0.01 is the initial value of learning rate, 'iter' is the current iteration, 'max\_iter' is the maximum iteration and 'power' = 0.5 is the coefficient that controls the reduction rate of learning rate after each iteration. The combination of using the 'learning rate decay' technique help to make the convergence faster, avoiding fluctuations around the optimal points.

## III. RESULTS

### A. Datasets

We evaluated the method on the part I in the ISIC Challenge 2017-Skin Lesion Analysis Towards Melanoma Detection. The training dataset consist of 2000 8-bit RGB dermoscopic images in various resolution and corresponding grayscale masks. The masks indicate the label for every pixel

in the dermoscopic images that the pixels belong to lesion area or not. The test dataset includes 600 similar pairs which are used for final evaluation. In the training phase, we shuffled the train dataset and split it in a 80-20 ratio, the use of the small piece, to be validation dataset. All of images and masks are resized to 144×192 resolution but maintain the aspect reasonable ratio between width and height like original images.

### B. Performance Evaluation

To evaluate the quantitative accuracy of segmentation results, we used the Dice similarity coefficient (DSC), and Jaccard (JAC) coefficient. The Dice coefficient measures the similarity between automatic and manual segmentations and is calculated as follows

$$DSC = \frac{2S_{am}}{S_a + S_m} \quad (1)$$

Jaccard coefficient (Jac) is also used to measure dissimilarity between two sets and defined as:

$$JAC = \frac{S_{am}}{S_a + S_m - S_{am}} \quad (2)$$

where  $S_a$ ,  $S_m$ , and  $S_{am}$  are, respectively, the automatically delineated region, the manually segmented region, and the intersection between the two regions.

### C. Results

To show segmentation performances of the proposed approach, we applied the proposed model to segment all images from the datasets. Some representative samples of the obtained results for such dataset are given in Fig. 2. In this figure, the original images, the segmentation by the proposed method are presented in the first and second row of the figure. The ground truths by human expert corresponding to images in the first row are also shown in the last row. As can be seen from this figure, by observation, there is a good agreement between the results obtained by the proposed approach and the ground truths.

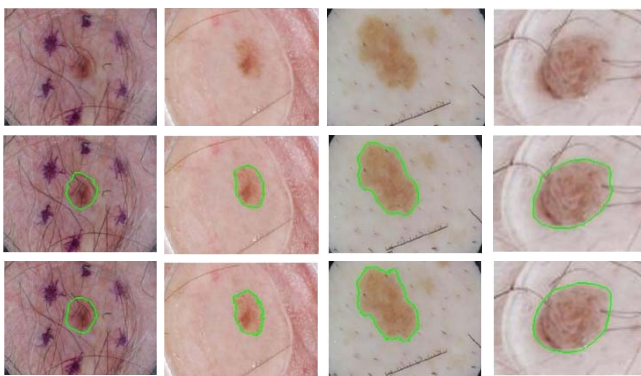


Figure 2. Representative segmentation results in the database by the Segnet approach. First row: input image, second row: result by the proposed modified SegNet, last row: ground truth

In order to show the convergence of the loss function while being minimized on both training and validation datasets as network training processes, we plot the learning curves of the proposed model when working on training set and the test set in Fig. 3a. In addition, the evolutions of DSC values vs epochs for the proposed model are also depicted in

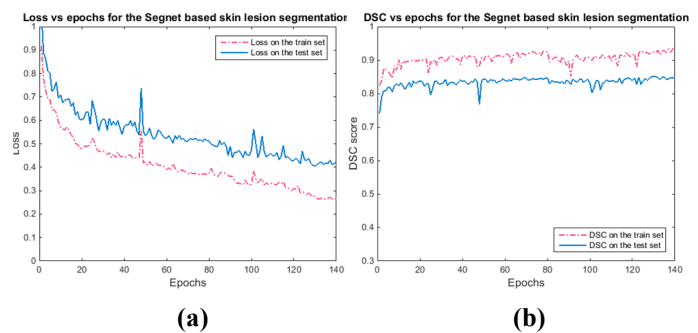


Fig.3. The Loss vs epochs (a), and The DSC score vs epochs (b) of the proposed approach when segmenting all images from the dataset

Fig. 3b. It can be seen from these figures; the proposed model converges when working with both training and validation datasets.

### D. Compared to other works

To quantitative evaluate the performances of the proposed model, we compare our model with other state-of-the-art methods included Galdran [17], Jahanifar et al. [18], Bi et al. [19], Xeu et al. [20], Ronneberger [21], FCN [14], and Yuan et al. [22] methods.

We report the average Dice Similarity Coefficient, and Jaccard coefficients of each method when segmenting all images from the database in Table 1. From Table 1, by quantitative comparison, we could see that the proposed model produces more accurate results than those by the comparative methods.

TABLE I. THE AVERAGE DICE SIMILARITY COEFFICIENT (DSC) AND JACCARD COEFFICIENT (JAC) BETWEEN OTHER STATE-OF-THE-ART AND THE PROPOSED MODELS ON THE ISIC CHALLENGE 2017.

Method	Dice Coefficient	Jaccard Coefficient
Galdran method [17]	0.810	0.718
Jahanifar et al. method [18] (2019)	0.827	0.721
Bi et al. method [19]	0.834	0.731
Xeu et al. method [20]	0.839	0.749
Ronneberger et al. method [21]	0.842	0.758
FCN method [14]	0.827	0.721
Yuan et al. method [22] (2019)	0.849	0.765
The proposed method	<b>0.853</b>	<b>0.771</b>

## IV. CONCLUSION

We have presented an approach for skin lesion segmentation. The paper has shown the effectiveness of the modification of SegNet deep learning networks for segmentation of skin lesion on dermoscopic images. The method can obtain the results with high accuracy as high as some other methods on multiple metrics. In addition, the method has other advantages in term of fast speed and not required manual prepared inputs. Furthermore, this method can be implemented in many different kinds of medical images segmentation. Along with the necessary modifications and

additions for the pipeline, the method could archive comparative performance as the state-of-arts in the future works.

#### ACKNOWLEDGMENT

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.05-2018.302.

#### REFERENCES

- [1] S. Pathan, K. G. Prabhu, and P. Siddalingaswamy, "Techniques and algorithms for computer aided diagnosis of pigmented skin lesions: A review," *Biomedical Signal Processing and Control*, vol. 39, pp. 237-262, 2018.
- [2] H. Rogers, M. Weinstock, S. Feldman, and B. Coldiron, "Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the U.S. Population, 2012," *JAMA Dermatology* vol. 151, pp. 1081-1086, 2015.
- [3] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA Cancer J. Clin.*, vol. 68, pp. 7-30, 2018.
- [4] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. Med. Imag.*, vol. 20, pp. 233-239, 2001.
- [5] F. Peruch, F. Bogo, M. Bonazza, V. Cappelleri, and E. Peserico, "Simpler, faster, more accurate melanocytic lesion segmentation through MEDS," *IEEE Trans Biomed Eng.*, vol. 61, 2014.
- [6] M. Celebi, H. Kingravi, H. Iyatomi, Y. Aslandogan, W. Stoecker, R. Moss, *et al.*, "Border detection in dermoscopy images using statistical region merging," *Ski. Res. Technol.*, vol. 14, pp. 347-353, 2008.
- [7] M. Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," *Skin Res. Technol.*, vol. 19, pp. e252-e258, 2013.
- [8] H. Zhou, X. Li, G. Schaefer, M. E. Celebi, and P. Miller, "Mean shift based gradient vector flow for image segmentation," *Comput. Vis. Image Understand.*, vol. 117, pp. 1004-1016, 2013.
- [9] A. Sadri, M. Zekri, S. Sadri, N. Gheissari, M. Mokhtari, and F. Kolahdouzan, "Segmentation of dermoscopy images using wavelet networks," *IEEE Trans Biomed Eng.*, vol. 60, pp. 1134-1141, 2013.
- [10] F. Xie and A. Bovik, "Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm," *Pattern Recognition*, vol. 46, pp. 1012-1019, 2013.
- [11] M. E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, and G. Schaefer, "A state-of-the-art survey on lesion border detection in dermoscopy images," in *Dermoscopy Image Analysis*, ed, 2015, pp. 97-129.
- [12] G. Venkatesh, Y. Naresh, S. Little, and N. O'Connor, "A Deep Residual Architecture for Skin Lesion Segmentation," in *OR 2.0 Context-Aware Operating Theaters Computer Assisted Robotic Endoscopy Clinical Image-Based Procedures and Skin Image Analysis*, Cham, Switzerland, 2018, pp. 277-284.
- [13] N. Ibtihaz and M. S. Rahman, "Multiresunet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," Available: <https://arxiv.org/abs/1902.04049>, 2019.
- [14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431-3440, 2015.
- [15] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis MRI," Available: <https://arxiv.org/abs/1604.00494>, 2016.
- [16] H. Ünver and E. Ayan, "Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm," *Diagnostics (Basel)*, vol. 9, p. E72, 2019.
- [17] A. Galdran, A. Alvarez-Gila, M. L. Meyer, C. L. Saratxaga, T. Araújo, E. Garrote, *et al.*, "Data-driven color augmentation techniques for deep skin image analysis," in *arXiv:1703.03702*, p. <https://arxiv.org/abs/1703.03702>, 2017.
- [18] M. Jahanifar, N. Z. Tajeddin, B. M. Asl, and A. Gooya, "Supervised saliency map driven segmentation of lesions in dermoscopic images," *IEEE J. Biomed. Health Inform.*, vol. 23, pp. 509-518, 2019.
- [19] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," in *arXiv:1703.04197*, p. Available: <https://arxiv.org/abs/1703.04197>, 2017.
- [20] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SegAN: Adversarial network with multi-scale L 1 loss for medical image segmentation," *Neuroinformatic*, vol. 16, 2018.
- [21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234-241.
- [22] Y. Yuan and Y.-C. Lo, "Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks," *IEEE J. Biomed. Health Inform.*, vol. 23, pp. 519-526, 2019.