Skin Lesion Segmentation Based on Improved U-net

Lina Liu*, Lichao Mou+, Xiao Xiang Zhu+, and Mrinal Mandal*

*University of Alberta,
lina1@ualberta.ca, mmandal@ualberta.ca
+German Aerospace Center, Technical University of Munich,
lichao.mou@dlr.de, xiaoxiang.zhu@dlr.de

***Abstract*— Melanoma is one of the most common and dangerous skin cancers, accounting for 75% of deaths associated with skin cancer. Detection of melanoma in early stages can significantly improve the survival rate. Automatic segmentation of melanoma is an important and essential step for accurate detection of melanoma. Many existing works based on traditional segmentation methods and deep learning methods have been proposed for high-resolution dermoscopy images. However, due to the intrinsic visual complexity and ambiguity among different skin conditions, automatic melanoma segmentation is still a challenging task for existing methods. Among these methods, the deep learning methods have obtained more attention recently due to its high performance by training an end-to-end framework, which needs no human interaction. U-net is a very popular deep learning model for medical image segmentation. In this paper, we propose an efficient skin lesion segmentation based on improved U-net model. Experiments conducted on the 2017 ISIC Challenge dataset towards melanoma detection shows that the proposed method can obtain state-of-the-art performance on skin lesion segmentation task.**

***Index Terms*— Skin Lesion Segmentation, U-net, Dilated Convolution**

I. INTRODUCTION

Skin cancer is the third most common human malignancy, among which, melanoma is the most aggressive kind of cancer. As per a recent statistic, the incidence of melanoma on the skin has risen rapidly over the last 30 years, and approximately 96,480 cases of melanoma will be newly diagnosed in the USA in 2019 [1]. Fortunately, an early detection of melanoma is highly curable before it spreads to the other body parts. The 5-year relative survival rate is 98% for localized stage. Dermoscopy is an imaging technique that eliminates the surface reflection of skin, with high resolution and enhanced visualization ability. It is used worldwide as a preliminary step to examine the suspected skin lesion, due to its non-invasive nature. However, because of the intrinsic visual similarity between different skin conditions, e.g., melanoma, nevus and seborrheic keratosis, it is very difficult to differentiate them even for the dermoscopy experts. A few examples from the 2017 ISIC Challenge dataset are shown in Fig. 1. Meanwhile, the availability of experienced dermatologists worldwide is highly limited and the decisions are biased among different dermatology experts. Thanks to the highly developed machine learning technology, some methods have obtained encouraging results which surpass human performance [2]. Consequently, it is possible and meaningful to design automated dermoscopic assessment algorithms to positively assist dermatologists and ameliorate the aforementioned difficulties.

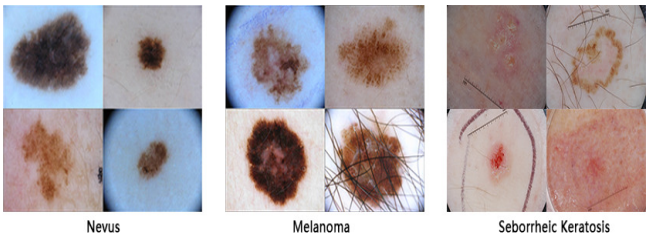

Fig. 1. Examples from 2017 ISIC challenge on skin lesion analysis towards melanoma detection dataset.

The general pipeline for the melanoma detection follows the steps of preprocessing, segmentation, feature extraction and classification. Skin lesion segmentation is a primary and essential step, which can benefit the subsequent classification task. In this paper, we will focus on the research of skin lesion segmentation task using dermoscopic images. The segmentation methods can be broadly divided into traditional methods and deep learning methods. The traditional methods are based on shallow learning models, which can be learned in a unsupervised or supervised manner, while the deep learning methods are based on the DCNN (deep convolution neural network) models, which can be learned in a supervised manner by using an end-to-end architecture. Many traditional methods using morphological operations along with the clustering techniques [3], [4] have been proposed. Jafari *et al.* [3] proposed to use k-means clustering to segment the skin lesion into foreground and background region. Similarly, Ali *et al.* [4] used Fuzzy C-means (FCM) to segment skin lesions. Another popular class of approaches is based on active contour models [5], [6], where the contour can evolve toward the boundaries of the interest regions. After getting the candidate regions using thresholding methods, active contour models driven by local histogram fitting energy [5] or multi-direction gradient vector flow (GVF) snake [6] could be used to refine the course segmentation. However, the traditional methods usually involve complex procedures and its performance is highly dependant on the pre-processing and post-processing steps. It will fail in cases where the skin condition is complex and the images are not ideal, as the

978-1-7281-0319-8/19/$31.00 ©2019 IEEE

Fig. 2. Schematic of the proposed technique.

examples shown in Fig. 1.

Recently, skin lesion segmentation using deep learning methods has become very popular. These methods are typically based on the DCNN models. The FCN (fully convolutional network) [7] proposed by Long *et al.* and U-net [8] proposed by Ronneberger *et al.* can be used for skin lesion segmentation. Jafari *et al.* [9] proposed a pixel-level segmentation method based on the FCN model. The output of the method is the prediction label for a pixel, and the inputs are image patches with different scales centered at this pixel so as to make use of the local and global context information. Yuan *et al.* [10] proposed a method based on the fully convolution-deconvolution method. Loss function based on Jaccard distance is used instead of the regular cross entropy loss. U -net is a very popular DCNN architecture because of its huge success on medical images with small data size. Some methods based on U-net [11], [12] for melanoma segmentation and classification have been proposed.

Although, much research about skin lesion segmentation has been done, it is still a challenging task due to the complex skin conditions and fuzzy boundaries. To address the aforementioned problem, we propose an improved method based on U-net architecture in this paper. Different from the original U-net, we introduce batch normalization layer and dilated convolution layer. The batch normalization layer is used to prevent the overfitting problem, and dilated convolution layer is used so as to enlarge the perceptive field during training but without losing resolution. Moreover, a simple test ensemble technique is used in this paper, which improve the performance significantly. Compared with existing ensemble techniques, which usually train multiple models using different parameters, our method does not require training extra models, thus being very efficient. Details of the proposed method can be found in section II.

II. Proposed Method

The proposed technique has three modules (see Fig. 2): pre-processing, DCNN based segmentation and test ensemble. In this section, we present details of the modules of the proposed technique. As the training data is limited, data augmentation is required to train the CNN networks. We also explain the data augmentation used in this paper.

A. Pre-processing

The original images are of high resolution 1022×767, and it is difficult to train a deep learning model using the high resolution images. To train the deep model more efficiently, in our method, all the input images are first augmented and then resized to the same scale 256×256. We used RGB and HSV color channels (i.e., 6 channels) as input since the RGB combined with HSV color space can provide more information than either one color space alone. All input channels are normalized to the range [0, 1] by dividing the maximum intensity values of each color channel.

B. DCNN based Segmentation

Schematic of the proposed DCNN model is shown in Fig. 3. The input consists of 6 channels corresponding to RGB and HSV spaces, and the output is the (segmentation) probability map. The number of feature maps are shown under each operation block. All filters used in this model are of size 3×3. As we can see from Fig. 3, the proposed method contains an encoder path and a decoder path, which is composed of a sequence of encoder blocks and decoder blocks, respectively. The encoder blocks use the following structure: *conv*, *BN*, *conv*, *BN*, *dilated_conv* and *max_pool*. The *conv* is the convolution layer, *BN* is the batch normalization layer, *dilated_conv* is the dilated convolution layer at stride of [2, 2], and *max_pool* is the max pooling layer with pool size [2, 2]. For each decoder block, upsampling layer of size [2, 2] is first used to increase the feature maps' resolution. We then concatenate these feature maps with feature maps of the same size from the encoder path (as the green arrows show in Fig. 3). Therefore, the overall structure of the decoder block is: *up_sampling*, *concatenate*, *conv*, *BN*, *conv*, *BN* and *dilated_conv*. The yellow block Conv_5 in Fig. 3, which is the connection layer for the encoder path and decoder path, is consisted of *conv*, *BN*, *conv*, *BN* and *dilated_conv*. In our method, for all the convolution layers, we set stride to be 1, same padding is used and Rectified Linear Units (ReLus) are used as the activation function. Compared with the original U-net architecture, we add batch normalization layer and dilated convolution layer in our model. The batch normalization layer performs normalization in each mini-batch during training, which can reduce the internal covariate shift of networks during training. It is useful to speedup the training procedure, and can be used as a regularization term to avoid the overfitting problem [13]. Dilated convolution [14] of rate 2 is used in our method. The dilated convolution can increase the perceptive field which is very suitable for skin lesion segmentation task, since the interest regions of dermoscopic images are of similar visual patterns but of different scales. Experiments in section III has verified that the introduction of dilated layer can significantly improve the performance of the original U-net model.

With successive convolutional layers and pooling layers, the encoding path can extract the contextual and semantic information along the encoding path, resulting in reduced resolution. The upsmapling layer performs the reverse operation of max pooling layer and increase the resolution of feature maps. By subsequent convolution operation, the final output is a probability map with each element representing the probability of a pixel belonging to the interest region (melanoma, nevus or seborrheic keratosis). The entire model can be learned in an end-to-end manner using the training data.

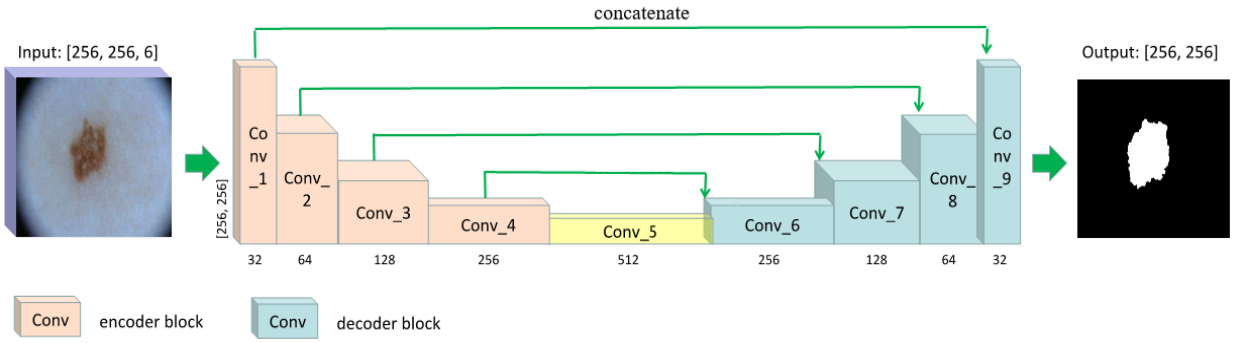

Fig. 3. Schematic of the proposed DCNN model. The proposed model consists of two pathways, encoding path and decoding path. The contextual and semantic information can be extracted along the encoding path by downsampling in each encoder block. Full image resolution mask can be obtained at the end of decoder path by the upsampling operation in each decoder block.

C. Test Ensemble

To improve the segmentation performance, many existing works get the final results by ensembling different trained models. The final probability map is the average mean value of probability maps generated using different learned models. However, training different models, especially the DCNN models can be very time-consuming. In this paper, we use a simple test ensemble technique by implementing data augmentation technique on the test data. For a given test image, we first get the transformed images by rotating the raw images 90, 180, 270 degree, flipping vertically and horizontally. After getting the predicted masks of these generated test images, we reconstruct the prediction of the raw image by reverse operation on the generated masks. The the final prediction mask is then calculated as the average mean value of these reconstruction maps.

D. Data Augmentation

The successful implementation of deep learning models requires large amount of training data. Data augmentation is used in this paper so as to increase the number of training samples from 2000 images to 48000 images. Due to the fact that the interest region of a skin lesion is usually centered in the image, and it usually occupies smaller area compared with the skin regions, we use images of 3 different scales. The 3 scaled images include the original images, and the cropped centered images with 80% and 70% of its original size. Then for image of each scale, we randomly rotate the image by 90, 180, 270 degree, or randomly flip the images. In addition, we also use rigid Moving Least Squares method [15] to generate deformation images. The generated images are of slight visual difference images compared with the original images, which can mimic the real images and are suitable for the medical image segmentation task. All these images are then normalized to the same size of 256×256.

III. Experiments and Results

A. Dataset

The dataset using in this project is ISIC 2017 challenge skin lesion analysis towards melanoma detection [16], which is a very challenging dataset. Resolution for the lesion image is 1022×767. There are 2000 images in the training set, including 374 melanoma, 254 seborrheic keratosis, and the remainder as benign nevi (1372). A separate validation dataset (150 images) could be used for participants to submit automated results for evaluation. The final test dataset contains 600 images. All the images are of high-resolution. Illumination variation, noise and various artifacts are also witnessed in this dataset.

B. Implementations

The proposed segmentation algorithm was implemented with Keras based on Tensorflow. All the experiments were conducted on a Dell Aliware desktop with Intel(R) i7-7700 4.2 GHz CPU and a GPU of Nvidia GeForce GTX 1080Ti with 11GB memory. In this method, we adopt Adam optimization algorithm for neural network training. The learning rate is set to be 0.0001, and cross entropy loss is used as the lost function. The number of epochs during training is set to be 3, and the batch size is set to be 16. Therefore, steps per epoch equals to $4800016 = 3000$.

C. Evaluation Metrics

In this paper, the Jaccard index (JA), Dice coefficient (DC), Accuracy (ACC), Sensitivity (SE) and Specificity(SP) are used to evaluate the proposed method, which are calculated as follows:

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

$$DC = \frac{2 \cdot TP}{2 \cdot TP+FP+FN} \tag{2}$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

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

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

where TP (True Positive) is the number of foreground pixels being correctly classified as foreground (interest region). TN

(True Negative) is the number of background pixels being correctly classified as background (skin region). FP (False Positive) is the number of background pixels being wrongly classified as foreground. FN (False Negative) is the number of foreground pixels being wrongly classified as background.

D. Results and Analysis

In this section, we compare the proposed method with the U-net architecture [8] and SegNet [17] architecture. In order to make a fair comparison, the input of these neural networks are the same. Experiment results can be found in Table I. Proposed_1 and Proposed_2 represent the proposed method without test ensemble and with test ensemble, respectively. As we can see from from Table I, the proposed method obtain the best performance, and the adoption of dilated convolution layer can significantly improve the performance. Meanwhile, Proposed_2 outperforms Proposed_1, which shows the importance of test data ensemble. Examples of the output probability maps can be found in Fig. 4.

TABLE I
COMPARISON OF DIFFERENT METHODS ON ISIC 2017 SKIN LESION SEGMENTATION DATASET.

Method	JA	DC	ACC	SE	SP
SegNet [17]	0.700	0.797	0.917	0.815	0.974
U-net [8]	0.686	0.788	0.915	0.755	0.969
Proposed_1	0.740	0.830	0.926	0.828	0.965
Proposed_2	0.752	0.840	0.930	0.829	0.988

IV. Conclusions and Future Work

In this paper, we propose an improved DCNN model based on U-net for skin lesion segmentation. Compared with the original U-net, our method adopted batch normalization layer to avoid the overfitting problem, and dilated convolution layer to increase the size of receptive field during training. Experiment results have shown that the introduction of dilated convolution can significantly improve the performance of proposed method. Meanwhile, we also introduce a simple yet useful test ensemble technique which does not require training extra models. During the testing phase, we have noticed that the proposed model is more likely to segment the dark regions as the interest regions, and it fails in a few cases where the interest regions are lighter than the surrounding skin regions. This is typically an imbalance data problem. In our future work, techniques that could tackle the class imbalance problem will be used to address the aforementioned problem.

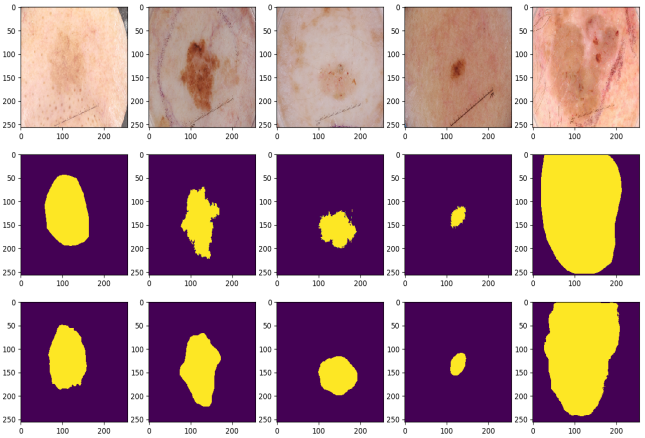

Fig. 4. Examples of our method's performance on 2017 ISIC challenge dataset for skin lesion segmentation. The first row contains the RGB images, the second row is the ground truth mask and the third row is the predictions of the proposed method.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a cancer journal for clinicians*, vol. 69, no. 1, pp. 7–34, 2019.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *IEEE international conference on computer vision*, Santiago, Chile, Dec. 2015, pp. 1026–1034.

[3] M. H. Jafari, S. Samavi, S. M. R. Soroushmehr, H. Mohaghegh, N. Karimi, and K. Najarian, "Set of descriptors for skin cancer diagnosis using non-dermoscopic color images," in *International Conference on Image Processing (ICIP)*, Arizona, USA, Sep. 2016, pp. 2638–2642.

[4] A.-R. Ali, M. S. Couceiro, and A. E. Hassenian, "Melanoma detection using fuzzy c-means clustering coupled with mathematical morphology," in *International Conference on Hybrid Intelligent Systems (HIS)*, Delhi, India, Dec. 2014, pp. 73–78.

[5] Q. Abbas, I. Fondón, A. Sarmiento, and M. E. Celebi, "An improved segmentation method for non-melanoma skin lesions using active contour model," in *International Conference Image Analysis and Recognition*, Algarve, Portugal, Oct. 2014, pp. 193–200.

[6] J. Tang, "A multi-direction gvf snake for the segmentation of skin cancer images," *Pattern Recognition*, vol. 42, no. 6, pp. 1172–1179, 2009.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE conference on computer vision and pattern recognition*, Florida, USA, Jun. 2015, pp. 3431–3440.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, Oct. 2015, pp. 234–241.

[9] M. H. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S. M. R. Soroushmehr, K. Ward, and K. Najarian, "Skin lesion segmentation in clinical images using deep learning," in *International Conference on Pattern Recognition (ICPR)*, Cancún, Mexico, Dec. 2016, pp. 337–342.

[10] Y. Yuan, "Automatic skin lesion segmentation with fully convolutional-deconvolutional networks," *arXiv preprint arXiv:1703.05165*, 2017.

[11] M. Berseth, "Isic 2017-skin lesion analysis towards melanoma detection," *arXiv preprint arXiv:1703.00523*, 2017.

[12] H. Chang, "Skin cancer reorganization and classification with deep neural network," *arXiv preprint arXiv:1703.00534*, 2017.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[14] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[15] S. Schaefer, T. McPhail, and J. Warren, "Image deformation using moving least squares," in *ACM transactions on graphics (TOG)*, vol. 25, no. 3, 2006, pp. 533–540.

[16] N. C. Codella, D. Gutman, and et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *International Symposium on Biomedical Imaging*, Washington, D.C., USA, Apr. 2018, pp. 168–172.

[17] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.