

SkinNet: A Deep Learning Framework for Skin Lesion Segmentation

Sulaiman Vesal, Nishant Ravikumar and Andreas Maier

Abstract—There has been a steady increase in the incidence of skin cancer worldwide, with a high rate of mortality. Early detection and segmentation of skin lesions is crucial for timely diagnosis and treatment, necessary to improve the survival rate of patients. However, skin lesion segmentation is a challenging task due to the low contrast of lesions and their high similarity in terms of appearance, to healthy tissue. This underlines the need for an accurate and automatic approach for skin lesion segmentation. To tackle this issue, we propose a convolutional neural network (CNN) called SkinNet. The proposed CNN is a modified version of U-Net. The network employs dilated and densely block convolutions to incorporate multi-scale and global context information during training. We compared the performance of our approach with other state-of-the-art techniques, using the ISBI 2017 challenge dataset. Our approach outperformed the others in terms of the Dice coefficient, Jaccard index and sensitivity, evaluated on the held-out challenge test data set, across 5-fold cross validation experiments. SkinNet achieved an average value of 85.10, 76.67 and 93%, for the DC, JI and SE, respectively.

I. INTRODUCTION

OVER 5 million new cases of skin cancer are diagnosed in the United States alone, each year. Melanoma is the advanced form of skin cancer and the global incidence of melanoma was estimated to be over 350,000 cases in 2015, with almost 60,000 deaths. Although the mortality rate is significant, early detection improves the survival rate to over 95%[1][2]. Computer-aided-diagnostic systems that enable automatic and accurate skin lesion detection, segmentation and classification are thus essential. However, automatic segmentation of lesion in dermoscopy images is challenging. This is due to significant variation in lesion tumor morphology (shape, size, structure, occluding hair), intrinsic image attributes (skin pigmentation, color, contrast) and acquisition parameters (imaging devices, resolution, lens deformation). Recently, deep learning methods based on convolutional neural network (CNN) showed significant performance in the field of semantic segmentation. CNN methods have the capability to extract more prominent features from the entire image rather than hand-crafted features. However, many studies also attempted to address skin lesion segmentation task using CNN, for example: [1] proposed a deep learning framework consisting of two fully convolutional residual networks, to simultaneously segment and classify skin lesions; [3] introduced a deformable model using a newly defined speed function

and stopping criterion for skin lesion segmentation; and [4] used a deep learning approach called fully convolutional residual network (FCRN) with more than 50 layers for both segmentation and classification. [5] developed a full resolution convolutional networks method to learn the full resolution features of each individual pixel of the input image without any further pre- or post-processing operations. A multi-stage CNNs proposed by [6] which is learning features from both the training data and the estimated results derived from the previous fully convolutional network stage.

In this study we propose a CNN-based skin lesion segmentation framework, called SkinNet. The proposed CNN architecture is a modified version of the U-Net [7]. The latter has demonstrated state-of-the-art performance in various medical image segmentation tasks in recent years. The U-net architecture basically consists of a contracting path (encoder), which downsamples an image into a set of high-level features, followed by a symmetric expanding path (decoder), which uses the feature information to build a pixel-wise segmentation mask. SkinNet employs dilated convolutions [8] in the lowest layer of the encoder-branch in the U-Net, to provide a more global context for the features extracted from the image. Additionally, we replace the conventional convolution layers in both the encoder and decoder branches of U-Net, with dense convolution blocks, to better incorporate multi-scale image information. In each dense block in both encoder/decoder parts, we have two convolutions + ReLU activation layers. The input feature map is concatenated with the output of next convolution layers. Section II describes the proposed method, dataset and experimental setup and the results are summarized in Section III.

II. METHODS

Incorporating both local and global information, is generally beneficial for any segmentation task. However, in a conventional U-Net, the lowest level of the network has a small receptive field which prevents the network from extracting features that capture non-local image information. Dilated convolutions [8] provide a suitable solution to this problem. They introduce an additional parameter, called the dilation rate, to convolution layers, which defines the spacing between values in a kernel. This helps dilate the kernel such that a 3×3 kernel with a dilation rate of 2 will have a receptive field size equal to that of a 5×5 kernel. Additionally, this is achieved without any increase in complexity, as the number of parameters associated with the kernel remains the same which prevents the network from extracting features that capture non-

S. Vesal, N.Ravikumar, and A. Maier are with the Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany (telephone: +49 152 10228542, e-mail: {sulaiman.vesal, nishant.kumar, andreas.maier}@fau.de).

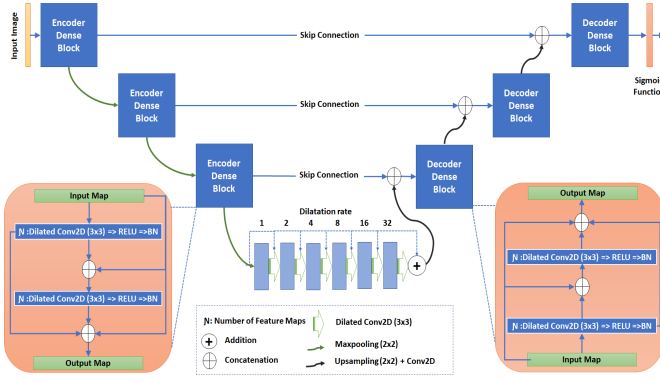


Fig. 1: SkinNet architecture: in the bottleneck layer feature maps are convolved with a dilatation rate of 1-32. The dense blocks concatenate the feature maps from the previous layer with the current output feature map.

local image information. Furthermore, to incorporate multi-scale image information in the network, we replaced normal convolution layers at each level of both the encoder and decoder-branches, with densely connected convolution layers. In the dense convolution blocks, the input to each layer is a concatenation of outputs (or feature maps) from all preceding convolution layers. The overall architecture of SkinNet is summarized in Fig.2. The dense blocks show in blue boxes comprise two convolution layers with a kernel size of 3×3 . At the bottleneck, encoded features are convolved using different dilation rates.

A. Pre-processing

Since most of images in the training dataset have various scales, all training images first normalized using min-max normalization technique and then resized to 512×512 pixels. We also performed data augmentation in order to boost the performance of SkinNet. The input images are randomly augmented using various image transformation techniques such as rotation, flipping, color shifting, translation and scaling operations. Fig.2. shows three images from training dataset with their random argumentations.

B. Loss Function

To measure the performance of the model, we defined a dice coefficient loss function which is summed over the classes:

$$\zeta(y, \hat{y}) = 1 - \sum_k \frac{\sum_n y_{nk} \hat{y}_{nk}}{\sum_n y_{nk} + \sum_n \hat{y}_{nk}} \quad (1)$$

\hat{y}_{nk} denotes the output of the model, where n runs over all pixels and k runs over the classes (in our case, background vs. skin lesion). The ground truth masks are one-hot encoded and denoted by y_{nk} . We take one minus the dice coefficient in order to constrain the loss toward zero.

C. Evaluation and Results

In order to evaluate the performance of SkinNet, we trained it on the ISBI 2017 challenge dataset [2], which includes

TABLE I: Lesion segmentation performances of different frameworks.

Method	AC	DC	JI	SE	SP
Yuan et al. [10]	0.934	0.849	0.765	0.825	0.975
Berseth et al. [2]	0.932	0.847	0.762	0.820	0.978
Attia et al. [11]	0.936	0.824	0.738	0.836	0.966
Li et al. [1]	0.952	0.839	0.753	0.855	0.974
SkinNet	0.932	0.851	0.767	0.930	0.905

2000 dermoscopic images and the corresponding lesion masks. The lesion masks are binary mask created by expert manual tracing of the lesion boundaries. The organizers also provided a validation dataset that includes 150 images and an additional test dataset with 600 images for final evaluation. Lesion segmentation accuracy of our network was evaluated with respect to provided ground truth masks, using two well known similarity metrics: the Dice similarity coefficient (DC) and the Jaccard index (JI). We also calculated the sensitivity (SE), specificity (SP) and Accuracy (AC) metrics [3] in order to directly compare our method with the state-of-the-art, presented in recent studies. We trained the network using the Adam optimizer [9] with a mini-batch size of 8. ADAM was chosen as it provides an elegant mechanism for adaptively changing the learning rate, based on the first and second-order moments of the gradient, at each iteration. The initial learning rate is set as 0.0001 and reduced during the training. The network trained with 5 folds cross validation on the training dataset and tested and validated with another 600 and 150 images. Fig.2. shows the average of training plot for 100 epochs and as it can be seen, there is no overfitting during the training process. The results for the test dataset are also shown in Table 1. respectively. It can be seen that SkinNet outperformed other methods in terms of DC, JI and sensitivity with 85.1%, 76.7% and 93.0%.

III. DISCUSSION AND CONCLUSION

While various network configurations have been proposed for skin lesion segmentation, to the best of our knowledge none of them employ an architecture identical to SkinNet. In this paper, we proposed a novel deep learning approach which exploits both local and global image information, for skin lesion segmentation. Our method achieved higher JI and DC compared to other state-of-the-art methods (as shown in Table 1) which used various types of CNN architectures. The dense blocks incorporate multiscale level information effectively in the encoder and decoder branches of the network. The JI and DC values achieved by our network were 76.67 and 85.10%, respectively. Additionally, SkinNet achieved the highest sensitivity (93%), outperforming the rest.

ACKNOWLEDGMENT

This study was partially supported by the project BIG-THERA: Integrative ‘Big Data Modeling’ for the development of novel therapeutic approaches for breast cancer, and the BMBF grant: Using big data to advance breast cancer research. The authors would also like to thank NVIDIA for donating a Titan XP GPU.

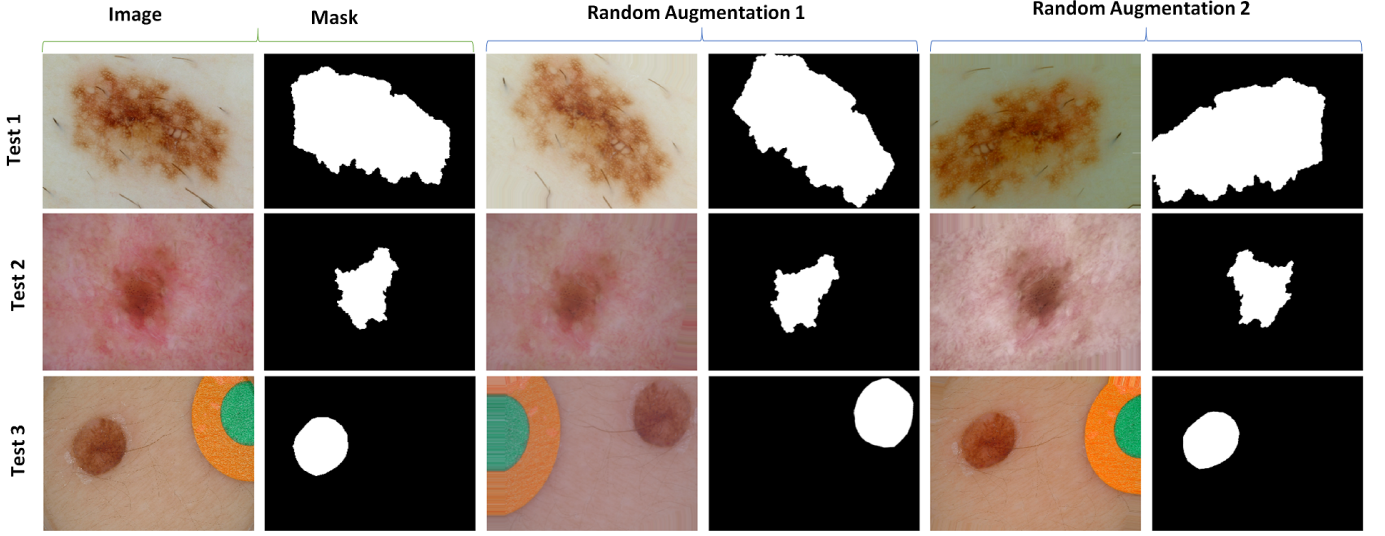


Fig. 2: Dermoscopy images augmentation. The first and second columns show the training images with their masks and columns 3-6 shows augmented images with various rotation, flipping, color shifting, translation and scaling operations.

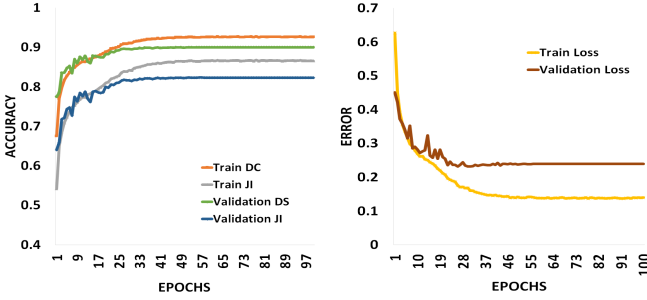


Fig. 3: Training and validation loss, DS and JI curves for 5 fold cross validation (average).

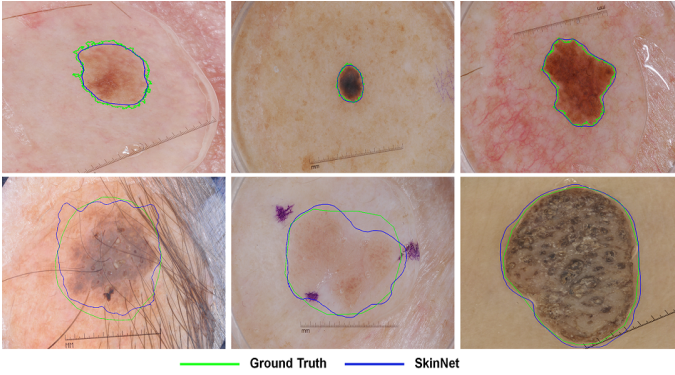


Fig. 4: SkinNet segmentation output on ISBI 2017 test dataset. The green contour shows the ground truth segmentation and blue SkinNet output respectively.

REFERENCES

- [1] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *Sensors*, vol. 18, no. 2, 2018.
- [2] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "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 *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, April 2018, pp. 168–172.
- [3] Z. Ma and J. M. R. S. Tavares, "A novel approach to segment skin lesions in dermoscopic images based on a deformable model," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 2, pp. 615–623, March 2016.
- [4] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, April 2017.
- [5] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 221 – 231, 2018.
- [6] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, "Dermoscopic image segmentation via multistage fully convolutional networks," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2065–2074, Sept 2017.
- [7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015, pp. 234–241.
- [8] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *CoRR*, vol. abs/1511.07122, 2015.
- [9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proceedings of the 3rd International Conference on Learning Representations*, vol. abs/1412.6980, 2014.
- [10] Y. Yuan, M. Chao, and Y. C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance," *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1876–1886, Sept 2017.
- [11] M. Attia, M. Hossny, S. Nahavandi, and A. Yazdabadi, "Spatially aware melanoma segmentation using hybrid deep learning techniques," *CoRR*, vol. abs/1702.07963, 2017.