A Deep Residual Architecture for Skin Lesion Segmentation

Venkatesh G M, Naresh Y G, Suzanne Little, and Noel E. O'Connor

Insight Centre for Data Analytics-DCU,
Dublin City University, Ireland
venkatesh.gurummunirathnam@insight-centre.org
naresh.yarlapati@insight-centre.org
suzanne.little@insight-centre.org, noel.oconnor@insight-centre.org

Abstract. In this paper, we propose an automatic approach to skin lesion region segmentation based on a deep learning architecture with multi-scale residual connections. The architecture of the proposed model is based on UNet [22] with residual connections to maximise the learning capability and performance of the network. The information lost in the encoder stages due to the max-pooling layer at each level is preserved through the multi-scale residual connections. To corroborate the efficacy of the proposed model, extensive experiments are conducted on the ISIC 2017 challenge dataset without using any external dermatologic image set. An extensive comparative analysis is presented with contemporary methodologies to highlight the promising performance of the proposed methodology.

Keywords: Skin Lesion · FCNs · Residual connection · U-Net.

1 Introduction

Medical imaging is an emerging and successful tool increasingly employed in precision medicine. It aids in making a medical decision for providing appropriate and optimal therapies to an individual patient. Skin Cancer is one such disease which can be identified through medical imaging using dermoscopic techniques. There are many types of skins cancers, but we can broadly put them in to two general categories viz., Non melanoma and Melanoma. Non-melanoma cancers are unlikely to spread to other parts of the body but Melanoma is likely to spread to other parts of the body and is known to be aggressive cancer. Malignant Melanoma is a cutaneous disease. It affects the melanin producing cells known as melanocytes. Melanoma is likely to be fatal, it has caused more deaths than any other type of skin disease [18]. The dermoscopic acquisition of a skin image targets segmentation into two regions: lesion and normal skin. The affected part of an organ or a tissue due to a disease or an injury is generally termed as lesion. Efficient and accurate segmentation of the lesion region in dermoscopic images aids in classification of various skin diseases. Furthermore, the severity of the diseases can be predicted through various grading techniques which result in

early identification of a skin disease which plays a vital role in the treatment and cure of the disease.

2   Literature

In the existing literature, several attempts have been made to develop a more robust and efficient segmentation of the lesion region in dermoscopic and non-dermoscopic clinical images. The methods for skin lesion segmentation can be segregated into the following categories [8] viz., thresholding, active contours, region merging methods [13] and deep learning architectures. Some methods [7, 16] have been proposed on non-dermoscopic images which address skin lesion segmentation based on colour features and textural properties respectively. The method in [15] addresses the illumination effects and artifacts. These methods apply post processing steps for refining the segmentation results. In [19], a deep convolutional neural network (CNN) has been proposed which combines both local texture and global structure information to predict a label for each pixel for segmentation of the lesion region. In [11], an automated system for skin lesion region segmentation has been proposed to classify each pixel based on pertinent geometrical, textural and colour features which are selected using Ant Colony Optimization (ACO). The complementary strengths of a saliency and Bayesian framework are applied to distinguish the shape and boundaries of the lesion region and background in [2]. In [23], an unsupervised methodology based on the wavelet lattice, shift and scale parameters of wavelets has been proposed for the segmentation of skin lesion regions in dermoscopic images. In [6], image-wise supervised learning is proposed to derive a probabilistic map for automated seed selection and multi-scale super-pixel based cellular automata to acquire structural information for skin lesion region segmentation. A Guassian membership function is applied for image fuzzification and to quantify each pixel for skin segmentation [12].

Despite several methods being available for segmentation of lesion region in images of skin diseases, there is still scope for exploring new models, which are efficient and provide better segmentation. Thus, in this work, we propose a deep residual architecture inspired by UNet [22] for skin lesion segmentation. The rest of this paper is organized as follows: Section 3 elaborates the proposed model for the segmentation of skin lesion region. Section 4 gives the experimental analysis and comparative analysis. Section 5 gives a conclusion.

3   Proposed Method

The proposed methodology for automatic skin lesion region segmentation using deep learning architecture is shown in Fig 1. The architecture is inspired from UNet [22] and residual network [17]. The input to the network is RGBH (Red, Green, Blue and Hue planes respectively) of a dermoscopic image and the output is binary segmented image with white and Black pixels representing the affected skin and non-affected regions respectively. There are four important components

in the proposed network. The first is construction of a multi-scale [14] image pyramid input which makes the network scale invariant. The second is a U-shape convolutional network, to learn a vivid hierarchical representation. The third is to incorporates residual learning to preserve spatial and contextual information from the preceding layers. The residual connections are used at two levels. Firstly, at each step of the contracting (encoder) and expansive (decoder) path of the U-Net and another short connection between the multi-scale input and expansive (encoder) path of respective step. The information lost in the encoder stages due to the max-pooling layer at each level is preserved through the multi-scale residual connection. Finally, a layer with binary cross-entropy loss function based on Jaccard index [3] is included for classification of pixels.

3.1   Multi-scale Input Layer

The proposed method has similar architecture to the methodology in [14] for constructing the multi-scale input by using an average pooling layer to downsample the images naturally and construct a multi-scale input in the encoder path. These scaled input layers are used to increase the network width of decoder path and also as a shortcut connection to the encoder path to increase the network width of the decoder path.

3.2   Network Structure

U-Net [22] is an efficient fully convolutional network which has been proposed for biomedical image segmentation. The proposed architecture adopts similar architecture consisting of two blocks placed in U-shape as shown in Fig 1. The block with green color (Fig2(a)) represents the residual downsampling block and the red color (Fig2(b)) represents residual upsampling block. A 2×2 max-pooling operation with stride 2 for downsampling is used and at each stage number of feature channels chosen in the proposed architecture are shown in Fig **??**. The left side path consist of repeated residual downsampling block (henceforth referred as resDownBlock) which are connected to the corresponding residual upsampling block (henceforth referred as resUpBlock). This connection is shown with dotted lines in Fig **??** similar to U-Net, where the feature maps of resDownBlock is concatenated to the corresponding resUpBlock. Along with the u-connection, there are also short connections between the Multi-scale input at each step of the U-Net with the corresponding resUpBlock by convolving the scaled input with 3×3 convolution which avoids convergence on a local optimal solution and thus helps the network to achieve good performance in complex image segmentation.

3.3   Residual-Down-sampling block(resDownBlock)

The structure of resDownBlock consists of two 3×3 convolutions, each followed by a rectified linear unit (ReLU). A shortcut connection of the input layer is added with the output feature-maps of the second convolution layer before passing to the ReLU as shown in Fig **??**. Batch Normalization is adopted between
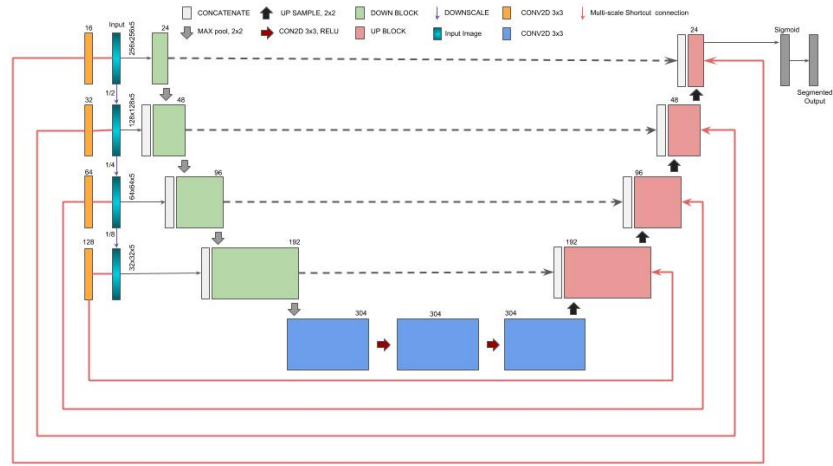
Fig. 1. Proposed architecture of Multi-Scale Residual UNet

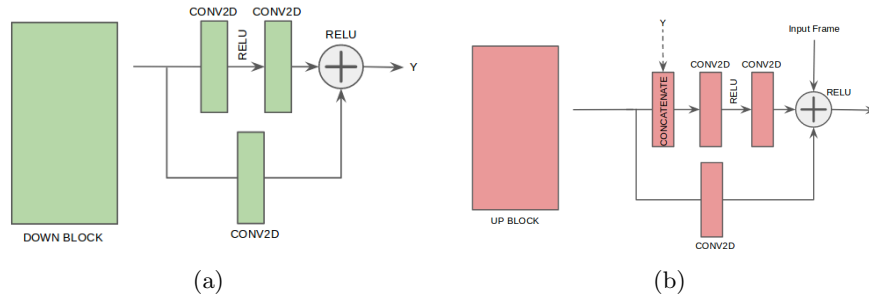
(a)   (b)

Fig. 2. (a) Details of resDownBlock, (b) Details of resUpBlock

convolutional layer and rectified linear units layer as well as during the shortcut connection. The max-pooling layer in the resDownBlock has a kernel size of 2×2 and a stride of 2. Excluding the initial resDownBlock in the encoder path all other resDownBLocks receives the concatenated output feature-maps from the preceding block with the scaled input.

3.4   Residual-Up-sampling block(resUpBlock)

The structure of resUpBlock consists of two 3×3 convolutional layers, each followed by a rectified linear unit (ReLU) and a shortcut connection of the input layer is added with the output feature-maps of second convolution layer along with the shortcut connection of the scaled input image before passing to the ReLU as shown in Fig **??**. There is a Concatenation layer, which concatenates the upsampled feature-maps from previous block with the feature-maps of resDownblock. According to the architecture, the resolution of resDownBlocks output should match with the resUpBlock's input for adding the upsampling

layer in the beginning of each block. Batch Normalization is adopted between convolutional layer and rectified linear units layer similarly as mentioned in the above section.

4   Experiments

In order to evealute the efficacy of the proposed model, experiments have been conducted on the **ISIC 2017 Challenge** [10] official dataset. The dataset consists of dermoscopic images with 2000 training, 150 validation and 600 test samples respectively. The proposed network is implemented using the Keras neural network API [9] with Tensorflow backend [1] and trained on a single GPU (GeForce GTX TITAN X, 12GB RAM). The network is optimized by Adam optimizer [20] with an initial learning rate of 0.001. For increasing the number of samples during the training phase, we have used standard geometrical (linear) data augmentation techniques, namely rotation(-45° to +45 °), horizontal and vertical flipping, translation and scaling(-10% to +10%)) of the input image. We choose 256×256 square images with batch size of 4 samples. The number of learning steps at each epoch is set to 1000. We have exploited RGB and HSI color space model for deriving RGBH (Red, Blue, Green and Hue Channels of dermoscopic images) as input data to the network to capture the color variations in the data. Fig. 3 presents the lesion region segmentation for few test samples with overlay of segmentation results. The overlay consists of differentiations viz., blue, green and red overlays representing false negatives, true positives and false positives respectively. It is evident that the proposed model effectively captures the lesion region without any post-processing steps.

To evaluate the performance of the segmentation, we have use Accuracy (AC), Jaccard Index (JA), Dice coefficient (DI), Sensitivity (SE) and Specificity (SP). Consider β_{tp}, β_{tn} , β_{fp} and β_{fn} which represent the number of true positive, true negative, false positive and false negative respectively. All the above mentioned metrics are computed using equations (1) - (5):

$$Accuracy(AC) = \frac{\beta_{tp} + \beta_{tn}}{\beta_{tp} + \beta_{tn} + \beta_{fp} + \beta_{fn}} \tag{1}$$

$$Sensitivity(SE) = \frac{\beta_{tp}}{\beta_{tp} + \beta_{fn}} \tag{2}$$

$$Dice\ coefficient(DI) = \frac{2 * \beta_{tp}}{2 * \beta_{tp} + \beta_{fp} + \beta_{fn}} \tag{3}$$

$$Specificity(SP) = \frac{\beta_{tn}}{\beta_{tp} + \beta_{fn}} \tag{4}$$

$$JaccardIndex(JA) = \frac{\beta_{tp}}{\beta_{tp} + \beta_{fp} + \beta_{fn}} \tag{5}$$

Fig. 4 presents the segmentation results of the proposed model compared to other methods in the literature by depicting number of test samples in each bin, where each bin in x-axis represents the Jaccard Index range, y-axis represents number of test samples. The results of our method is presented in Table. 1. From
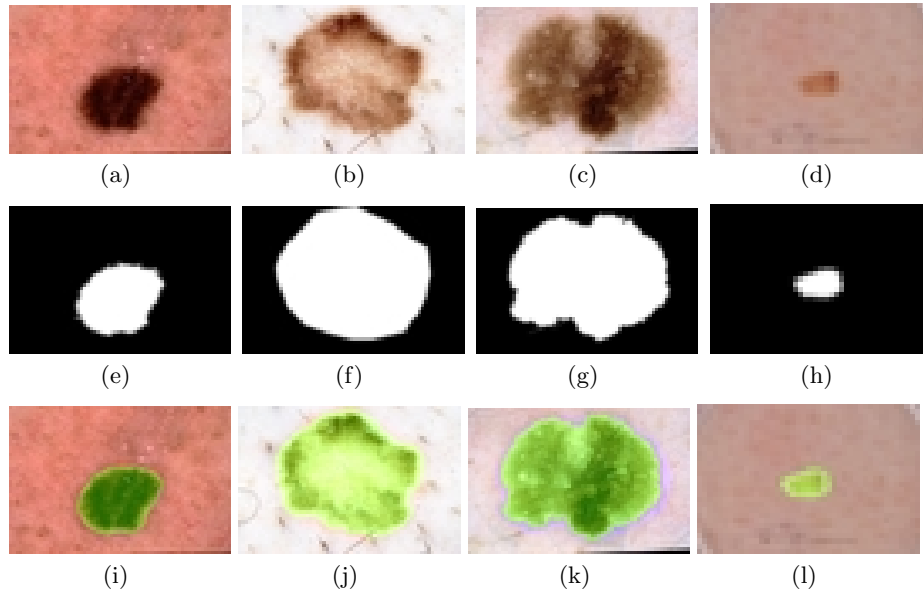
Fig. 3. The visual examples of lesion region segmentations, 1st row are test images, 2nd row corresponding ground truth images and 3rd are output of the segmented lession region

Table. 1, it is evident that the proposed method outperforms the other methods in terms of Accuracy, Dice Coefficient and Sensitivity. The results of our method is quite competitive for the ISIC 2017 dataset in comparison with the methods which have shown top performance in the literature.

Table 1. Comparison of Skin Lesion Segmentation on ISIC 2017.

Method	Accuracy	Dice Co-efficient	Jaccard Index	Sensitivity	Specificity
Yading Yuan [24]	0.934	0.849	**0.765**	0.825	0.975
Our Method	**0.936**	**0.856**	0.764	**0.83**	0.976
Matt Berseth [4]	0.932	0.847	0.762	0.82	0.978
popleyi [5]	0.934	0.844	0.76	0.802	**0.985**
Euijoon Ahn [5]	0.934	0.842	0.758	0.801	0.984
RECOD Titans [21]	0.931	0.839	0.754	0.817	0.97

5   Conclusion

In this work, we have proposed a deep architecture for skin lesion segmentation termed as Multi-scale residual UNet. From the results in Fig. 3, it can be observed that the boundaries of lesion regions and the background are well separated and differentiable. Furthermore, the proposed model uses only ≈16M parameters when compared to other well known conventional deep architectures for various complex applications. To further improve the performance, in our
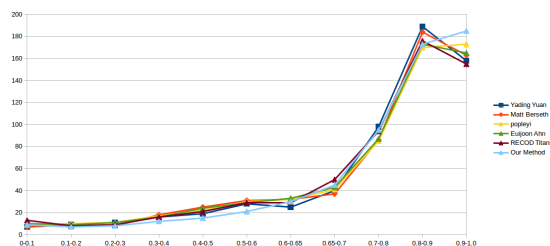
Fig. 4. Graphical representation of Jaccard Index on overall test set.

future work visual saliency shall be explored in conjunction with deep features and post processing methods based on Conditional Random fields.

Acknowledgement

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al.: Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467 (2016)
2. Ahn, E., Kim, J., Bi, L., Kumar, A., Li, C., Fulham, M., Feng, D.D.: Saliency-based lesion segmentation via background detection in dermoscopic images. IEEE journal of biomedical and health informatics **21**(6), 1685–1693 (2017)
3. Berman, M., Rannen Triki, A., Blaschko, M.B.: The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4413–4421 (2018)
4. Berseth, M.: Isic 2017-skin lesion analysis towards melanoma detection. arXiv preprint arXiv:1703.00523 (2017)
5. Bi, L., Kim, J., Ahn, E., Feng, D.: Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. arXiv preprint arXiv:1703.04197 (2017)
6. Bi, L., Kim, J., Ahn, E., Feng, D., Fulham, M.: Automated skin lesion segmentation via image-wise supervised learning and multi-scale superpixel based cellular automata. In: Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. pp. 1059–1062. IEEE (2016)
7. Cavalcanti, P.G., Yari, Y., Scharcanski, J.: Pigmented skin lesion segmentation on macroscopic images. In: Image and Vision Computing New Zealand (IVCNZ), 2010 25th International Conference of. pp. 1–7. IEEE (2010)
8. Celebi, M.E., Iyatomi, H., Schaefer, G., Stoecker, W.V.: Lesion border detection in dermoscopy images. Computerized medical imaging and graphics **33**(2), 148–153 (2009)
9. Chollet, F.: Keras https://github. com/fchollet/keras (2017)

10. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., 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: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. pp. 168–172. IEEE (2018)
11. Dalila, F., Zohra, A., Reda, K., Hocine, C.: Segmentation and classification of melanoma and benign skin lesions. Optik-International Journal for Light and Electron Optics **140**, 749–761 (2017)
12. Diniz, J.B., Cordeiro, F.R.: Automatic segmentation of melanoma in dermoscopy images using fuzzy numbers. In: 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS). pp. 150–155. IEEE (2017)
13. Emre Celebi, M., Kingravi, H.A., Iyatomi, H., Alp Aslandogan, Y., Stoecker, W.V., Moss, R.H., Malters, J.M., Grichnik, J.M., Marghoob, A.A., Rabinovitz, H.S., et al.: Border detection in dermoscopy images using statistical region merging. Skin Research and Technology **14**(3), 347–353 (2008)
14. Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X.: Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Transactions on Medical Imaging (2018)
15. Glaister, J., Amelard, R., Wong, A., Clausi, D.A.: Msim: Multistage illumination modeling of dermatological photographs for illumination-corrected skin lesion analysis. IEEE Transactions on Biomedical Engineering **60**(7), 1873–1883 (2013)
16. Glaister, J., Wong, A., Clausi, D.A.: Segmentation of skin lesions from digital images using joint statistical texture distinctiveness. IEEE transactions on biomedical engineering **61**(4), 1220–1230 (2014)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
18. Heath, M., Jaimes, N., Lemos, B., Mostaghimi, A., Wang, L.C., Peñas, P.F., Nghiem, P.: Clinical characteristics of merkel cell carcinoma at diagnosis in 195 patients: the aeiou features. Journal of the American Academy of Dermatology **58**(3), 375–381 (2008)
19. Jafari, M.H., Karimi, N., Nasr-Esfahani, E., Samavi, S., Soroushmehr, S.M.R., Ward, K., Najarian, K.: Skin lesion segmentation in clinical images using deep learning. In: Pattern Recognition (ICPR), 2016 23rd International Conference on. pp. 337–342. IEEE (2016)
20. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
21. Menegola, A., Tavares, J., Fornaciali, M., Li, L.T., Avila, S., Valle, E.: Recod titans at isic challenge 2017. arXiv preprint arXiv:1703.04819 (2017)
22. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
23. Sadri, A.R., Zekri, M., Sadri, S., Gheissari, N., Mokhtari, M., Kolahdouzan, F.: Segmentation of dermoscopy images using wavelet networks. IEEE Transactions on Biomedical Engineering **60**(4), 1134–1141 (2013)
24. Yuan, Y., Chao, M., Lo, Y.C.: Automatic skin lesion segmentation with fully convolutional-deconvolutional networks. arXiv preprint arXiv:1703.05165 (2017)