A Deep Fully Convolutional Network for Segmentation of Dermoscopic Images

Ananditha Raghunath[a], Shima Nofallah[b]

[a] *New York University Courant Institute of Mathematical Sciences*
[b] *University of Washington Department of Electrical Engineering*

Abstract

We present an automatic segmentation methodology for fast and accurate segmentation of dermoscopic images based on fully convolutional network architectures. Our algorithm was submitted to the ISIC 2018 Skin Challenge and the raw Jaccard index was 0.86 on the provided validation data.

---

1. Introduction

There are over 5 million new cases of skin cancer in the United States every year [1]. Although melanomas represent fewer than 5% of all skin cancers in the United States, they account for approximately 75% of all skin-cancer-related deaths. Every year they are responsible for over 10,000 deaths in the United States, and over 60,000 deaths globally [2, 3]. Observable as pigmented lesions occurring on the surface of the skin, melanomas can be detected by expert visual inspection. They are also amenable to automated detection with image analysis. Given the widespread availability of high-resolution cameras, algorithms that can improve the ability to screen and detect troublesome lesions can be of great value [3].

Segmentation, the partitioning an image into non-overlapping regions, each of which is homogeneous in one or more features and maximal in terms of this homogeneity, is usually the first process in the automated detection of melanoma. It brings to light regions of interest in the dermoscopic image that can be further examined. This paper explores a segmentation methodology based on Deep Learning, specifically Fully Convolutional Network (FCN) architectures, that have traditionally achieved high accuracy in computer vision related task such as segmentation, classification, etc [4].

2. Methodology

2.1. Data

Our data was extracted from the ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection grand challenge datasets [5, 6]. The training data consists of 2594 images and 2594 corresponding ground truth response masks of varying

Preprint submitted to Elsevier *July 28, 2018*

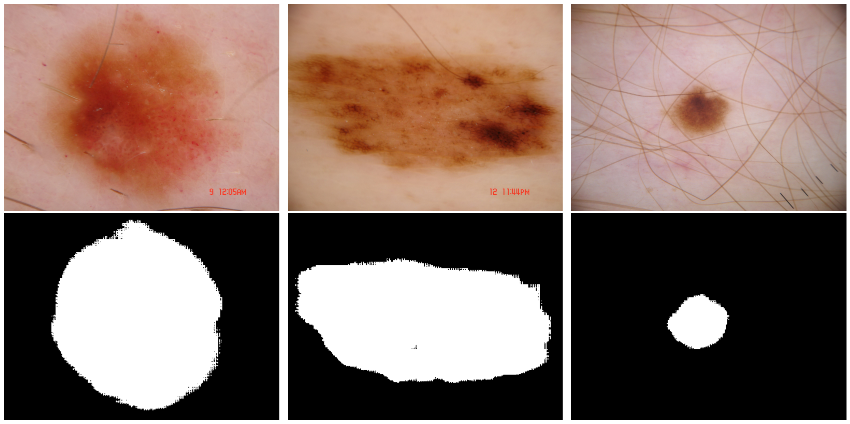

Figure 1: Sample images and associated segmentation predictions

sizes. The data was augmented using random rotations, lighting alterations, and random dihedral flipping so as to increase the training sample. Though most of the images were in a 3:4 aspect ratio, images were resized to squares when used for training to comply with the model's required input size. The model was trained at 2 input resolutions, 512*512 and 1024*1024, so as to obtain features maps of different sizes that could be used in tandem to segment.

2.2. Network

First attempts at segmentation were done without a pretrained network, and these didn't achieve compelling results. Furthermore, training took days and utlized large amounts of storage (over 10 GB). Fully convolutional networks (FCNs) extract features through the application of convolutional filters and train parameters through back propagation. As these have proved incredibly effective in representation learning, we use ResNet34 pretrained on the ImageNet dataset as an encoder [7]. To this backbone, cropped at the adaptive pooling layer, we finetuned using our data and attached a custom head to upsample the images. As information may be lost in the compressing of the images to filters and finally segmenting, the idea of Unet inspired activation copying and concatenation was used to decode [8]. A mixture of simple 2D Convolutions and cross convolutions were applied and concatenated at each level to output a probability map for each input image. These probability maps outputs of size 1024*1024 were then converted back to their original size using nearest neighbor interpolation and were subsequently outputted as final segmentation masks.

Code was written in Python with the Torch and Fast.ai libraries, and run on an Nvidia GPU. The learning rates and number of epochs were varied between the epochs of training (those of size 512*512 and those of size 1024*1024) to prevent overfitting on the training sample while finetuning. The learning rate for smaller images were $1e^{-2}$, $1e^{-3}$ and $1e^{-4}$ and for larger images the rates were $5e^{-2}$, $5e^{-3}$ and $5e^{-4}$. Adam optimizer was used to improve training in conjunction with BCEWithLogitsLoss which applied a sigmoid transform before calculating cross entropy loss. The total number of parameters trained exceeded 1 million.

3. Results

The accuracy on a random validation split of 500 images from the training sample as calculated by raw Jaccard index (intersection over union) was 0.86 on average. For the ISIC challenge, images with less than 0.65 Jaccard score were to be discarded when computing the final score. With this in mind, our resulting accuracy was 0.71 on the validation sample provided by the organizers.

4. Discussion

On examination of the images that the algorithm had difficulty with, we could see that modifications should be made when segmenting lesions that cover the entire image. This class of image was the only one which fell below 0.65 during validation. As scope for improvement, experimenting with different encoders such as Densenet or Resnet50 may yield better masks. Training at higher resolution, such as 2048*2048, may provide features that improve the borders of the masks. Due to storage limitations, these methods were not implemented. Increasing the sample even by 400 images in the final stage of image training improved performance significantly, thus with more training images the network will achieve greater accuracy.

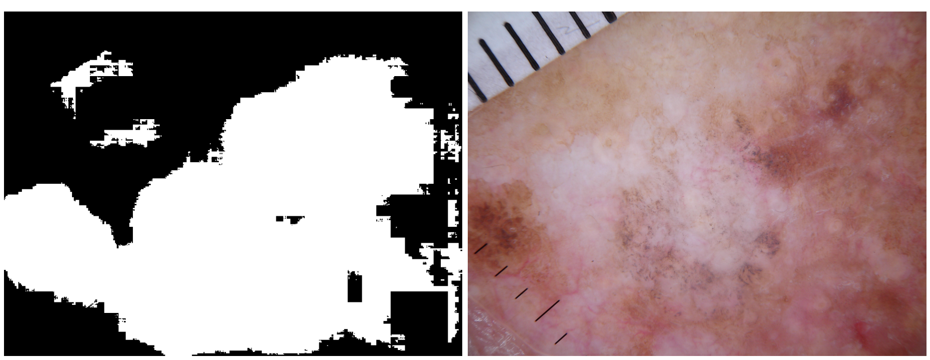

Figure 2: An image our algorithm couldn't accurately segment

5. Acknowledgements

This project was completed during a CRA-W DREU at the University of Washington. The authors appreciate the support of their advisor and mentor Dr. Linda Shapiro.

The methodologies used were inspired by information learned in Fast.ai and Stanford CS231n MOOCs. The authors would like to thank the educators for providing thorough online resources for one and all to access.

6. References

[1] H. W. Rogers, M. A. Weinstock, S. R. Feldman, B. M. Coldiron, Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the us population, 2012, JAMA dermatology 151 (2015) 1081–1086.

[2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (2017) 115.

[3] Skin lesion analysis towards melanoma detection, ISIC 2018 (2018).

[4] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.

[5] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, 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, IEEE, pp. 168–172.

[6] P. Tschandl, C. Rosendahl, H. Kittler, The ham10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions, arXiv preprint arXiv:1803.10417 (2018).

[7] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

[8] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, pp. 234–241.