High Resolution Human Skin Image Segmentation by means of Fully Convolutional Neural Networks

Manuel Calderon-Delgado[1†], Jeng-Wei Tjiu[2], Ming-Yi Lin[2] and Sheng-Lung Huang[1‡]
[1]Graduate Institute of Photonics and Optoelectronics, National Taiwan University, Taipei 10617, Taiwan
[2]Department of Dermatology, National Taiwan University Hospital and College of Medicine, Taipei 100, Taiwan
[†]mancaldel@ntu.edu.tw, [‡]shuang@ntu.edu.tw

***Abstract*—Convolutional neural networks were applied to sub-micron-resolution optical coherence tomography images of the human skin for anatomical segmentation. The main layers of skin were discerned with an average 90% accuracy, which we believe to possess potential in the assessment of skin health.**

Keywords—optical coherence tomography, machine learning, convolutional neural networks, skin, medical imaging

I. Introduction

The morphology of the dermal-epidermal junction (DEJ) has been proven to serve as an indicator of ageing [1] and cancer staging [2, 3]. Hence, if a machine learning algorithm is able to detect this region, it is highly probable that it will also discern between healthy and cancerous tissue or assess on the state of the skin for different pathologies.

Medical imaging techniques present a trade-off between resolution and imaging depth. Micrometer resolution has been achieved by different techniques, such as confocal microscopy and electron microscopy, but typical imaging depths are less than 100 µm and axial resolution is poor [4]. These techniques also require the excision of the tissue, which can result in complications due to scarring, bleeding and infections. Optical coherence tomography (OCT) offers an imaging depth of several hundred micrometers, while having sub-micron resolution. Moreover, it is able to perform *in vivo* measurements, reducing diagnose time and enabling physicians to perform multiple measurements *in situ*.

Machine learning has been highly successful in many computer vision applications [5], most of them using convolutional neural networks (CNNs). This type of neural network requires fewer variables than traditional "fully-connected" neural networks, which makes it possible to process larger images.

We have designed a fully convolutional network (FCN) capable of segmenting OCT images of human skin into different layers, which can help diagnose of skin diseases.

II. Methodology

A. Imaging system and ground truth acquisition

The training images were obtained using our home-made Ce:YAG Mirau-based full-field OCT system [6], which delivers an isotropic resolution of 0.9µm down to a depth of 200 µm, showing both papillary dermis and epidermis. Doctors were given four 3D tomograms, each with 439 images, to delineate the DEJ as a 5-µm band using ImageJ software [7]. The remaining layers (dermis, epidermis, glass and glycerol) were segmented by semi-automated macros and revised by experienced researchers in our laboratory.

B. Design of the neural network and training configuration

We used a customized implementation of U-Net [8], implemented in Keras [9] with TensorFlow backend [10], adapting the input and output sizes, as well as different hyperparameters, as shown in Fig.1. Categorical cross-entropy is used as the loss function, where different weights were assigned to each layer, compensating their differences in total area. This way, the loss corresponding to pixels in the DEJ layers is 20 times higher than dermis and epidermis, while glycerol and glass pixels are 10 times higher.

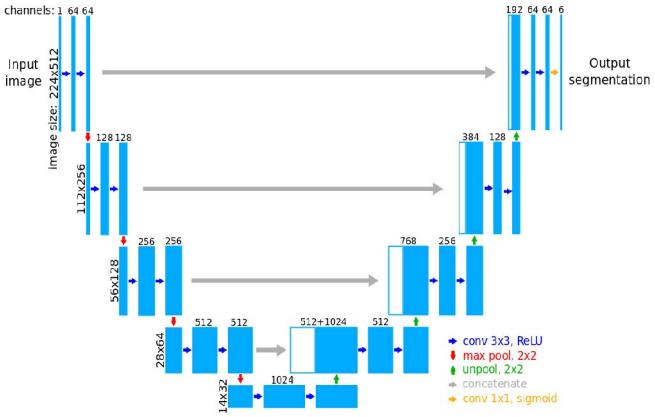

Fig.1: Schematic of the U-Net used in the simulations

The training was performed on the Taiwan GPU Cloud Center using 1 GPU (NVIDIA Tesla V100) and 8 CPUs (Intel Xeon E5) for 900 epochs, which took approximately 30 hours of computing time. The training set is composed of 3073 images of size 224×512 pixels from 4 different tomograms, while the test set consists of 439 similar images.

III. Results

We have obtained an average accuracy of 90% for both training and testing sets. The training was stopped because of the noise of the validation accuracy, probably due to the few segmented tomograms available.

Fig. 2 shows the normalized confusion matrix for the test set. The glass and glycerol classes have 100% and 96% accuracy, respectively, similar to human performance, while dermis (88%), epidermis (90%) and DEJ (91%) also show a high accuracy.

IV. Discussion

Fig. 3 shows the original OCT images, the doctors' segmentations and the predicted ones performed by our U-Net implementation. We have selected the most relevant ones to comment on them.

Most of the predicted results present perfect glass segmentation and it is nearly perfect for glycerol. It is remarkable that none of the bright spots on the epidermis have been segmented as glass, which induces us to think that the system is not only relying on the intensity values. Dermis and epidermis are also well placed by our network. The network has "learned" that epidermis lies right below the glass and glycerol layers.

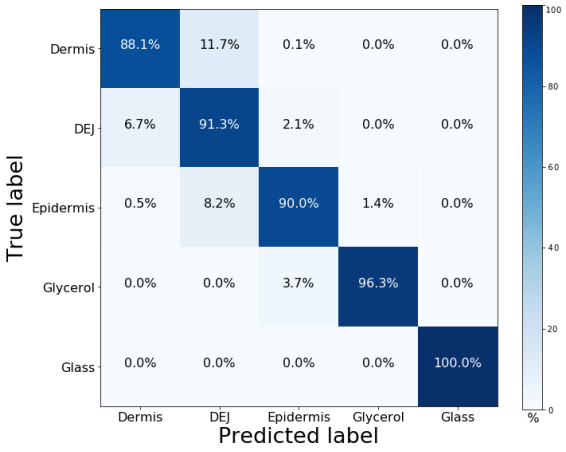

Fig 2: Normalized confusion matrix for the test set respect to each of the layers.

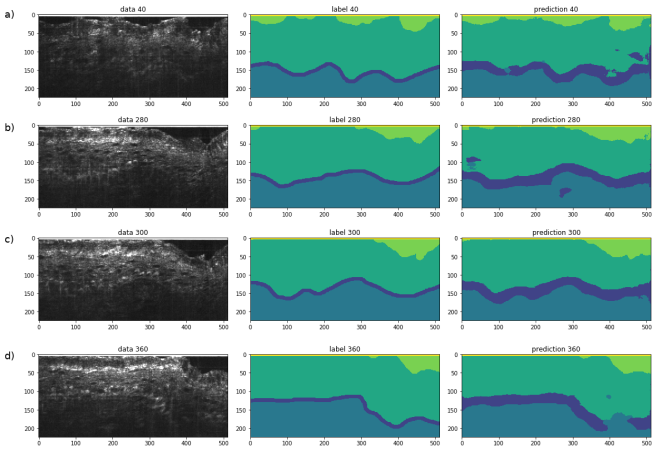

Fig. 3: Input OCT images of the skin, doctors' segmentations and predictions. Segmentation layers: dermis (blue), DEJ (dark blue), epidermis (green), glycerol (light green) and glass (yellow).

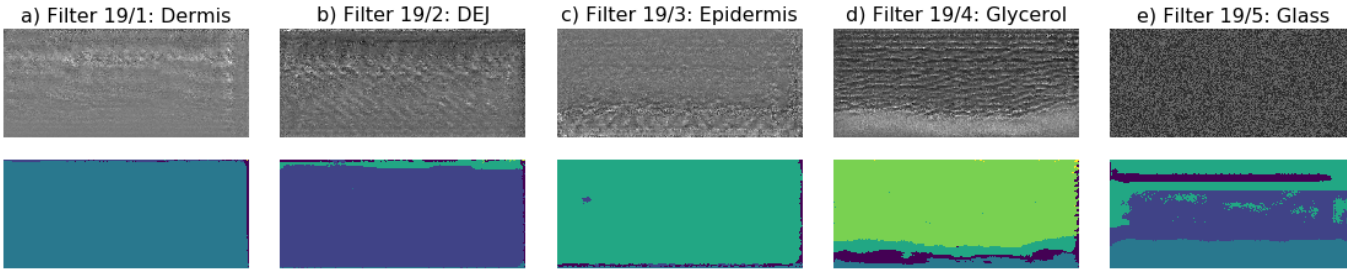

Figure 4: Input images maximizing last layer's output in the network (top) and their segmentation by the network (bottom).

The DEJ is, for most of the images, conformal to the doctors' segmentations. However, it is approximately double as thick. This is a consequence of the high weight given to this layer with respect to the other two: when the network tries to minimize the loss function, the misclassification penalty for a pixel is higher when the right label is DEJ as when it is dermis or epidermis, so by creating a thicker band, the average result will have a smaller loss function. In fact, in the first epochs of the training, the DEJ covered, approximately, one quarter of the image height, but it is then thinned as the network refines the filters' weights.

Another interesting feature is the DEJ island-like structures in Figs. 3.a and 3.b. There are two main reasons for this result. Firstly, our OCT images present complex structures that have not been segmented, such as Merkel or Langerhans cells, hair follicles and sweat pores, which present a high contrast change and might be confused by the system as the basal layer. Secondly, some of the areas in the edges present very low intensity values, making it difficult for the doctors to segment those areas with confidence.

Other problems arise in low intensity areas, where the predicted DEJ present discontinuities, as shown in Fig. 3.d. The system has not learned this property of skin, however, we expect to overcome this issue by training for a longer time, adjusting the network hyperparameters and including more training examples.

Lastly, it is interesting to observe the role assigned to the filters in the last layers of the network. Here we have six filters that can be better understood by finding the images that maximize their outputs. This was done by inputting random images and applying gradient ascent [11]. Fig. 4 shows the input images that maximize the last five filters of the network. We can observe a random pattern at specific locations of each image, suggesting that the network will not look for such layers at these areas. In fact, when we look at the prediction of the generated images (Fig. 4, bottom row) we see that these regions coincide with the segmentation assigned by the network to a wrong label. As expected, dermis and DEJ are not observed at shallow depths, while epidermis and glycerol are not found at deep regions.

However, despite the high accuracy obtained by the glass segmentation, the result for its associated filter does not have a clear interpretation (Fig.4.e). The generated input image looks noisy and its segmentation does not return a glass label. Nevertheless, we can observe a horizontal layered pattern with the expected order: dermis, DEJ and epidermis, from bottom to top.

V. Conclusions

We have demonstrated the feasibility of skin layers' segmentation by an FCN. Several issues are still to be resolved in order to increase the accuracy. However, we are still in a development phase where we are trying different network parameters in order to find an optimal configuration.

Acknowledgement: this research is supported by the R.O.C. Ministry of Science and Technology [MOST 107-2634-F-002-017].

References

[1] V. L. Newton, et al. Skin Res. Technol., 2017, 23: 131-148.
[2] F. Farnetani, et al. JAMA Dermatol. 2015, 151(10): 1075–1080.
[3] D. Rigel and A. Farberg. Elsevier, 1st Ed. 2017.
[4] M. R. Hamblin et al. Imaging in Dermatology, Academic Press, 2016.
[5] A. Voulodimos, et al. Computational Intelligence and Neuroscience, 2018.
[6] C. C. Tsai, et al. Biomed. Opt. Express 5, 2014: 3001-3010.
[7] J. Schindelin, et al. Nature methods 2012, 9(7): 676-682.
[8] O. Ronneberger, et al. MICCAI 2015, LNCS-9351: 234-241.
[9] F. Chollet, et al. "Keras," 2015 [https://keras.io].
[10] M. Abadi et al. Google 2015. [https://www.tensorflow.org].
[11] A. Mordvintsev, C. Olah and M. Tyka. Google AI Blog, June 2015.