Automated Detection of Noninvasive Imaging of Basal Cell Carcinoma by Convolutional Neural Network

Hsin-Fu Huang
National Taiwan University
hsinfu@cmlab.csie.ntu.edu.tw

Wei-Hung Weng
Harvard Medical School
ckbjimmy@hms.harvard.edu

Winston H. Hsu
National Taiwan University
whsu@ntu.edu.tw

Abstract

Diagnosis of basal cell carcinoma (BCC), the most common skin cancer, is made by histologic examination traditionally. Yet the process is invasive and time-consuming. In vivo imaging modalities such as harmonic generation microscopy (HGM) was therefore developed for noninvasive diagnosis of BCC. However the images acquired by HGM are too many for physicians to interpret manually. Thus, in this paper we focus on detecting features of BCC automatically by customizing compact and efficient convolutional neural network (CNN) models on HGM images of BCC. Our best model achieves a better result than AlexNet [1], while using less than its 1% number of parameters. The study indicated the potential solution of using customized CNN to detect the features in similar imaging modalities.

1 Introduction

Early diagnosis and intervention of basal cell carcinoma (BCC), the most common skin cancer, can reduce the complication of skin destruction, and decrease the risks of recurrence, spreading, and other skin cancers. Current diagnostic standard of BCC is histologic examination. However, the process is invasive and time-consuming. Therefore several noninvasive skin imaging techniques [2, 3, 4, 5, 6] such as harmonic generation microscopy (HGM) were developed to minimize unnecessary inconvenience.

In vivo noninvasive HGM was used to differentiate BCC from other nonmelanoma pigmented skin tumors morphologically [7]. The studies also used HGM to identify intratumoral melanocytes based on the nature of melanin as a source of HGM contrast [6, 8]. Yet the interpretation of noninvasive microscopic imaging is still a heavy load for physicians due to tons of image frames. Here we try to demonstrate the possible automated solution for identifying the morphological characteristics of BCC by using convolutional neural networks (CNNs).

CNNs are the current state-of-the-art method on the large scale natural image classification and detection competition. For example, ImageNet Large Scale Visual Recognition Challenge (ILSVRC). In the latest ILSVRC 2014, most of the CNN models [9, 10] were designed to be deeper, in order to better recognize complicated patterns on natural images. However, the patterns are usually much simpler on pathological images. For example, Cruz-Roa et al. [11, 12, 13] used CNNs on BCC histopathological images with only two convolutional layers (conv-layers) and one fully connected layer (fc-layer). To the best of our knowledge, no previous work discussed how to design a better CNN model for medical images, where image content and image sizes are very different from natural images. In this paper, we focus on customizing compact and efficient CNN models on our dataset and providing a guideline for future studies.

Table 1: Comparison of performance and parameter number of models

Model	Accuracy	Num of param
2 conv_8x8 layers + 1 fc_128 layer [12]	82.3%	conv-layers 36K
2 conv_3x3 layers + 1 fc_128 layer	82.0%	conv-layers 5K
3 conv_3x3 layers + 1 fc_128 layer	83.2%	conv-layers 14K
AleNet [1]	84.0%	all layers $>$ 60,000K
4 conv_3x3 layers + 2 fc_128 layers	84.4%	all layers 105K
5 conv_3x3 layers + 2 fc_128 layers	83.6%	all layers 65K

2 Dataset

The experimental images contain two channels, THG and SHG from HGM [7], with image size 515x512. There are 1133, 200 and 388 images with 1750, 223 and 802 labeled BCC-positive bounding boxes in the training, validation, and testing dataset. As most of the detection pipelines, we convert it to a classification problem by using sliding windows with fixed window size 64x64 and stride size 32x32. If the ratio of the BCC-positive area of a window is more than 0.8, we take this window as BCC-positive, and otherwise as BCC-negative. Due to the number of negative windows being much greater than the positive ones, we subsample 10% and 3% of the negative windows in training and validation set respectively, in order to balance the ratio to allow training the CNN model. However, we evaluate average precision (AP) instead of subsampling windows on our testing set, since windows cannot be skipped in practical applications. Currently, we have 8131, 921 and 2779 BCC-positive windows with 22778, 1236 and 75137 BCC-negative windows in the training, validation and testing dataset.

3 Methodology and Experiments

The philosophy of our model design, originated from VGG-Net [9], uses multiple conv-layers with small kernels to approximate the layer with large kernel. This decreases the number of parameters to avoid overfitting. For example, using two conv-layers with 3x3 kernel can cover the all possibilities of one conv-layer 5x5 kernel, and the former only takes 2x(3x3) = 18 parameters compared with the latter 5x5 = 25. The top three rows of Table 1 show the performance gain and model efficiency from small kernels.

Based on the philosophy of small kernels, we use 3x3 kernel size for every conv-layer, followed by a 2x2 max-pooling-layer. The number of kernels of the first conv-layer is 16 while the rest is 32. The number of nodes of the fc-layers is 128. We did the experiments with the combination of 0 5 conv-layers and 1 2 fc-layers. The last two rows of Table 1 show that the best accuracy is 4 conv-layers with 2 fc-layers instead of 5 conv-layers with 2 fc-layers, which means that the BCC patterns of our dataset are much simpler than natural images, the latter needs 16 conv-layers and 2 fc-layers to get the best results [9]. The AP of our best model on the testing set is 0.409.

4 Discussion and Conclusion

Here are two main reasons that we need to design our own model. First, our image content, from SHG and THG, is different than natural images, so using pre-trained models on natural images to fine-tune our dataset is unreasonable due to a big difference in image pattern basis. Second, the input image size of AlexNet is 256x256 but our image size is 64x64. Although we could resize 64x64 to 256x256 to fit in AlexNet [1], the last three rows of Table 1 show it is an overkill for our dataset.

This paper provides a guideline of customizing a compact and efficient model for medical images by using deeper conv-layers with small 3x3 kernels. Table 1 shows that this philosophy of small kernels can not only work better than large kernels but also achieve better performance with much less parameters than AlexNet [1].

References

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[2] Milind Rajadhyaksha, Melanie Grossman, Dina Esterowitz, Robert H Webb, and R Rox Anderson. In vivo confocal scanning laser microscopy of human skin: melanin provides strong contrast. *Journal of Investigative Dermatology*, 104(6):946–952, 1995.

[3] KJ Busam, C Charles, CM Lohmann, A Marghoob, M Goldgeier, and AC Halpern. Detection of intraepidermal malignant melanoma in vivo by confocal scanning laser microscopy. *Melanoma research*, 12(4):349–355, 2002.

[4] Mihaela Balu, Kristen M Kelly, Christopher B Zachary, Ronald M Harris, Tatiana B Krasieva, Karsten König, Anthony J Durkin, and Bruce J Tromberg. Distinguishing between benign and malignant melanocytic nevi by in vivo multiphoton microscopy. *Cancer research*, 74(10):2688–2697, 2014.

[5] Szu-Yu Chen, Hai-Yin Wu, and Chi-Kuang Sun. In vivo harmonic generation biopsy of human skin. *Journal of biomedical optics*, 14(6):060505–060505, 2009.

[6] Szu-Yu Chen, Shee-Uan Chen, Hai-Yin Wu, Wen-Jeng Lee, Yi-Hua Liao, and Chi-Kuang Sun. In vivo virtual biopsy of human skin by using noninvasive higher harmonic generation microscopy. *Selected Topics in Quantum Electronics, IEEE Journal of*, 16(3):478–492, 2010.

[7] Ming-Rung Tsai, Yu-Hsiang Cheng, Jau-Shiuh Chen, Yi-Shuan Sheen, Yi-Hua Liao, and Chi-Kuang Sun. Differential diagnosis of nonmelanoma pigmented skin lesions based on harmonic generation microscopy. *Journal of biomedical optics*, 19(3):036001–036001, 2014.

[8] Wei-Hung Weng, Ming-Rung Tsai, Yi-Hua Liao, and Chi-Kuang Sun. Differentiating pigmented skin tumors by the tumor-associated melanocytes based on in vivo third harmonic generation microscopy. In *SPIE Photonics West BIOS 2015 Technical Summaries*, page 3, 2015.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[10] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*, 2014.

[11] Angel Alfonso Cruz-Roa, John Edison Arevalo Ovalle, Anant Madabhushi, and Fabio Augusto González Osorio. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 403–410. Springer, 2013.

[12] Angel Cruz-Roa, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, and Anant Madabhushi. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In *SPIE Medical Imaging*, pages 904103–904103. International Society for Optics and Photonics, 2014.

[13] Haibo Wang, Angel Cruz-Roa, Ajay Basavanhally, Hannah Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio Gonzalez, and Anant Madabhushi. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, 1(3):034003–034003, 2014.