Skin Lesion Analysis by Multi-Target Deep Neural Networks

Xulei Yang[1], *Senior Member, IEEE*, Hangxing Li[2], Li Wang[1], Si Yong Yeo[3], Yi Su[3], Zeng Zeng[1]

Abstract—**Automatic skin lesion analysis involves two critical steps: lesion segmentation and lesion classification. In this work, we propose a novel multi-target deep convolutional neural network (DCNN) to simultaneously tackle the problem of segmentation and classification. Based on U-Net and GoogleNet, a single model is constructed with three different targets of both lesion segmentation and two independent binary lesion classifications (i.e., melanoma detection and seborrheic keratosis identification), aiming to explore the differences and commonalities over different target models. We conduct experiments on dermoscopic images from the International Skin Imaging Collaboration (ISIC) 2017 Challenge. Results of our multi-target DCNN model demonstrates superiority over single model with one target only (such as U-net or GoogleNet), indicating its learning efficiency and potential for application in automatic skin lesion diagnosis. To the best of our knowledge, this work is the first demonstration for a single end-to-end deep neural network model that simultaneously handle both segmentation and classification in the field of skin lesion analysis.**

I. Introduction

Skin cancer is a major public health problem and it is estimated that one in five Americans may be affected by the disease [1]. Melanoma is a common type of skin cancer where accurate diagnosis of skin lesion is crucial for its early detection.

Dermoscopic evaluation of melanocytic lesions (or moles) to detect melanoma is the current standard in clinical practice. Dermatologists specialised in managing skin diseases see patients for concerns about new or changing moles and lesions. In such evaluation, specific features are to be detected under the dermoscope and these features are tabulated using validated algorithms to help determine the risk of melanoma [2] by the dermatologists. However, the multitude of dermoscopic features are tedious to detect and the algorithms involved are complex and overwhelming [3]. As such, many dermatologists failed to utilize dermoscopic tools properly or accurately, which may compromise outcome of the evaluation.

Different imaging techniques, such as multispectral imaging and confocal microscopy, have been implemented to address the issues encountered in the detection of melanomas. However, such imaging devices are costly and bulky, and the dermatologist has to be trained in these imaging modalities. It has been shown that dermoscopic examination by trained and experienced dermatologists yields better sensitivity and specificity [4] in skin lesion diagnosis. Therefore, an automatic technique for robust analysis of dermoscopic dataset can be advantageous to clinicians.

Advanced dermoscopic algorithms [5, 6], such as "chaos and clues", "3-point checklist", "ABCD rule", "Menzies method", "7-point checklist" and "CASH" had been developed to facilitate the novice's ability to distinguish melanomas from benign with high diagnostic accuracy. Among these clinical evaluation algorithms, studies have shown that pattern analysis yields better diagnostic performance over other approaches [7]. The performance of pattern analysis for melanoma detection, however, greatly depends on the choice of meaningful descriptive features derived from the dermoscopic images. Identification of such features requires domain-specific expert knowledge and may fail for complex image segmentation and classification problems.

Deep neural networks techniques [8], in particular, the deep convolutional neural networks (DCNN) [9], have dramatically improved the state-of-the-art in object categorization and object detection. The DCNN has also been widely used on biomedical dataset, such as for skin lesion analysis [10, 11]. Specific to the ISIC 2017 Challenge on skin lesion analysis [12], the top results were achieved by teams that deployed deep neural networks to ensure precise boundary detection and accurate melanoma classification [13, 14]. Notably in DCNN, different features detected at the different convolutional layers allow the network to handle large variations in the dataset, obtain classification rules and handle feature detection automatically, thus ameliorating the difficulties of feature detection inherent in conventional pattern analysis techniques. Although various DCNN-based methods [10-14] have been proposed for skin lesion analysis, their approach involve treating the segmentation and classification process separately as single individual target. In other words, these approaches do not involve an end-to-end network that handles the multi-target problem of segmentation and classification simultaneously.

Inspired by the research work in [15], we propose a novel end-to-end multi-target DCNN to automatically solve both segmentation and classification of skin lesions at the same time. Different from conventional DCNN, such network architecture associates multiple labels that describe different characteristics of lesion with input dermoscopic images. Figure 1 shows the overall diagram of the proposed multi-target DCNN for skin lesion analysis, where three branches are designed based on the U-Net [16] and GoogleNet [17]

[1]X.L. Yang, L. Wang, Z. Zeng are with Institute for Infocomm Research, A*STAR, 1 Fusionopolis Way, Singapore 138632.

[2]H.X. Li is with National University of Singapore, 21 Lower Kent Ridge Rd, Singapore 119077. e-mail: hangxing.li@u.nus.edu

[3]Si Yong Yeo, Yi Su are with Institute of High Performance Computing, A*STAR, 1 Fusionopolis Way, Singapore 138632.

*corresponding authors, Xulei Yang/Zeng Zeng, phone: 6408-2299/2144; e-mail: yangxl, zengz@i2r.a-star.edu.sg

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

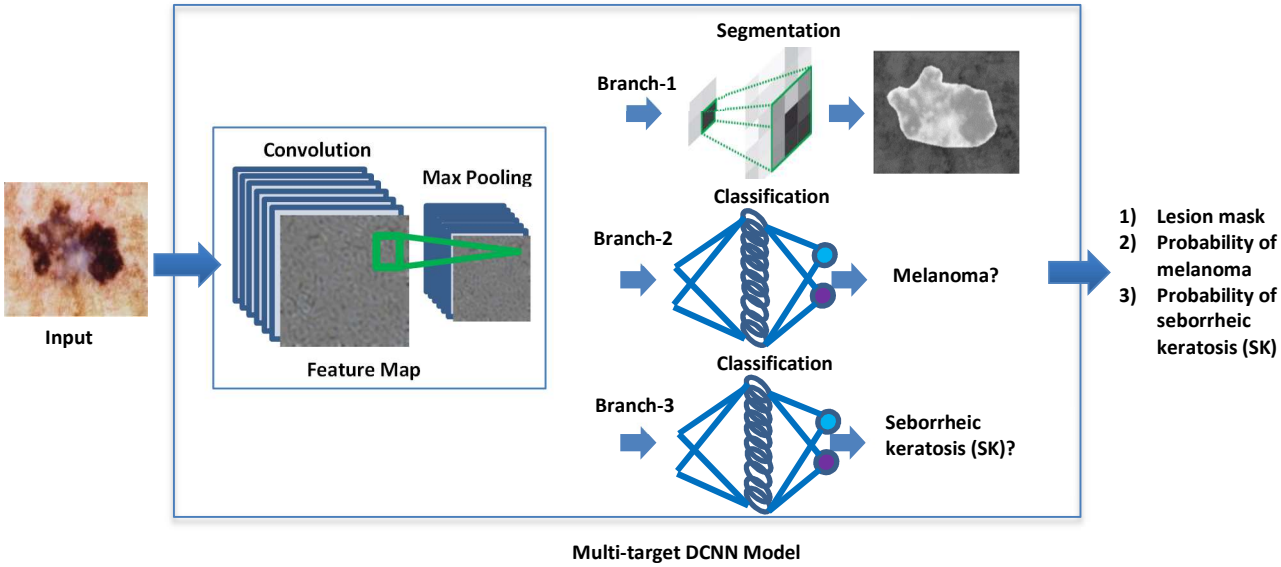

Fig. 1 Diagram of the proposed multi-target DCNN model for skin lesion analysis.

architecture (refer to next Section for details), with branch one targeting at lesion segmentation, branch two and three aiming at detecting melanoma and seborrheic keratosis (SK) separately. The de-convolutional layers and fully-connected layers are both constructed sequentially for segmentation and classification. This approach eliminates the need for hand-coded features as well as prior knowledge of the data, and can further handle even more targets due to its stability and scalability [15].

The contributions of this paper are three folds: First, to the best of our knowledge, this work is the first implementation of end-to-end networks to solve multi-target problems of both segmentation and classification in the domain of skin lesion detection. Next, rigorous experiments are conducted to demonstrate the superior performance of multi-target DCNN models over single model like the U-Net and GoogleNet. Finally, the preliminary results indicate the hypothesis that the multi-target network branches may boost each other to reach a better solution.

II. The Proposed Multi-Target DCNN Model

A. Network Architecture

The GoogleNet is broadly known as the winner of the Large Scale Visual Recognition Challenge 2014 (ILSVRC14), and its architecture involve a 22-layer deep neural network to solve single target classification [17]; the U-Net replaces the max-pooling layer with convolutional layers to generate semantic segmentation of medical images with very limited dataset [16]. Drawing from the GoogleNet and U-Net, the architecture of our proposed multi-target DCNN model is as shown in Figure 2. Five convolutional layers together with max-pooling layers (from GoogleNet architecture) are used to extract different layers of features which are shared for all three targets. Branch-1 includes five additional deconvolutional / upsampling layers (from U-Net architecture) targeting at object segmentation, while Branch-2 and Branch-3 uses one additional average-pooling layer and two fully-connected layers targeting at classification of melanoma and SK separately. Nonlinear activation function, namely the rectified linear unit (ReLU), follows every convolution and deconvolution operation; softmax and sigmoid functions served as the activation functions in the output operation for segmentation and classification, respectively.

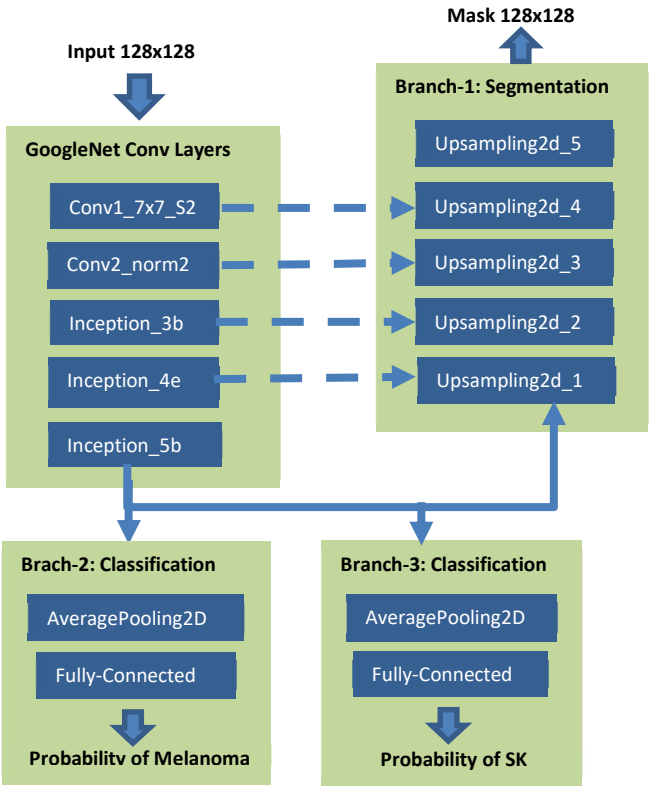

Fig. 2 Architecture of the proposed multi-target DCNN model

B. Loss Function

The proposed multi-target DCNN model contains one input dermoscopic image and three target models which

trained on dataset annotated by clinicians. The annotation includes both categorization of the skin lesion and the delineation of the lesion region. With regards to each target, the first branch is tasked to generate a binary mask image, wherein the lesion object within the image is indicated; the second branch aims to classify the image into two categories, i.e., melanoma or non-melanoma; the last branch targets at classifying the image into SK or non-SK categories. Since multiple branches are involved for different targets, each target model and their corresponding loss function are illustrated as following:

The First Target Model: This branch aims to detect the region of skin lesion. Given a region A representing an automatically segmented object and a region G representing a manually segmented object (ground truth), we follow the implementation in [18] to define the loss function based on the Dice Coefficient (DC) to maximize the overlap between A and G as follows:.

$$loss = 1 - DC = 1 - \frac{2(A \cap G)}{A + G} \quad (1)$$

The Second and Third Target models: Both models can be regarded as a binary classifier with different targets. As such, we deploy binary cross entropy to maximize the accuracy of the classification and the equation is illustrated as following.

$$\log loss = \frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (2)$$

where n is the number of images, y_i the ground truth of the category: 1 if the image is melanoma (or SK) for the second (or third) model, and 0 otherwise. The variable $\hat{y}_i$ is the predicted probability of the image being melanoma (or SK) for the second (or third) model.

III. Experiments and Discussions

A. Dataset and Experimental Setting

The dermoscopic images from the International Skin Imaging Collaboration (ISIC) 2017 Challenge on "skin lesion analysis toward melanoma detection" is employed for experiments [13]. The dataset contains 2000 training samples (374 melanoma and 254 SK) with annotations for both lesion classification and manually-traced lesion borders by clinicians, 150 dermoscopic dataset for validation and 600 for testing. The ISIC 2017 contains three tasks in competition, including skin lesion segmentation, localization of dermoscopic features and lesion classification. The last target on lesion classification basically includes two individual classification problems which, in our proposed network architecture, will be handled by Branch 2 and Branch 3 as aforementioned. In this work, we focus only on the lesion segmentation and classification tasks, and defer the feature localization task as future work.

B. Data Preprocessing

Deep neural networks need large amount of training data to achieve good performance. To build a powerful model using limited training data, image augmentation is very important to boost the performance of deep networks. The combination of multiple image processing steps, e.g., random rotation, shifts, shear, scale and flips, were implemented to create artificial images. We also implemented mean-variance normalization, color space transformation, and elastic transformation to enhance the augmentation.

C. Evaluation Metrics

The segmentation results are evaluated using the Jaccard Index, also known as Intersection-over-Union (IoU). The IoU is a measure of overlap between the area of the automatically segmented region and that of the manually segmented region. The value of IoU ranges from 0 to 1, with a higher value implying better match between the two regions.

The classification results are evaluated by the area under the receiver operating characteristic curve (AUC). The AUC is a measure of how well a parameter can distinguish between two diagnostic groups diseased/normal). The value of AUC ranges from 0 to 1, and larger AUC demonstrates superior classification outcome over others.

D. Training

The deep learning models were constructed on the Keras deep learning platform. The Adam optimizer with learning rate = 0.01, beta_1 = 0.9, beta_2 = 0.999, and epsilon = 1e-08 were used for compiling the Keras model. All weights were initialized with 0 mean and 0.1 standard deviation Gaussian distribution. These weights are optimized by back propagation through the aforementioned loss functions to calculate the penalty between the prediction and the ground truth in every batch. The network was trained for 15 epochs with a mini batch size of 32 using a 10-fold cross validation. The final accuracy is estimated by averaging the 10 different values produced by each fold.

E. Numerical Results

The multi-target network is compared against conventional DCNN, e.g., GoogleNet for lesion disease classification and U-Net for lesion segmentation. In this study, a 10-fold cross-validation is used to train and evaluate all the deep learning models. The trained deep learning models are tested on 150 evaluation and 600 testing samples, and the results are submitted to the challenge website for numerical assessment based on the aforementioned evaluation metrics.

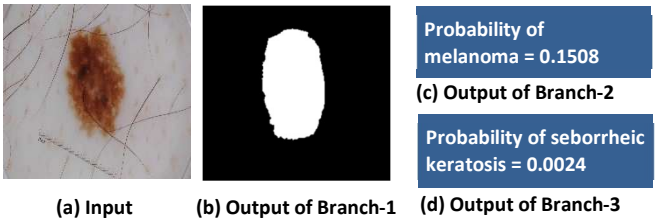

Fig.3 Output of our proposed multi-target DCNN model on skin lesion image ISIC_0012684 [13].

Qualitatively, Figure 3 illustrates the output of our proposed multi-target model on evaluation sample "ISIC_0012484" (Fig. 3a). The model simultaneously

produces three results including segmentation (Fig. 3b), probability of melanoma (Fig. 3c), and probability of seborrheic keratosis (Fig. 3d).

Quantitatively, Table 1 summarizes the average values of the segmentation evaluation metric IoU achieved by our proposed multi-target model and the well-known U-Net model [16]. Table 2 summarizes the average values of classification evaluation metric AUC achieved by the proposed multi-target model and the well-known GoogleNet model [17].

Table.1 Performance comparison in terms of IoU between proposed multi-target model and U-Net on ISIC 2017 Challenge task - lesion segmentation

Method	Our proposed model	U-Net [16]
Validation Set	0.776	0.764
Testing Set	0.741	0.737

Table 2 Performance comparison in term of AUC between proposed multi-target model and GoogleNet on ISIC 2017 Challenge Task - disease classification

Method	Our proposed model	GoogleNet [17]
Validation Set	0.926	0.903
Testing Set	0.886	0.857

F. Discussions

From the comparisons in Table 1 and Table 2, it was observed that the proposed multi-target deep learning model achieves better segmentation and classification performance as compared to individual U-Net and GoogleNet models, respectively. The segmentation appears to boost the classification performance quite significantly, but the converse seems not apply. Nevertheless, it has to be highlighted that the proposed model integrates multiple targets into one single model to achieve much higher training and testing efficiency in terms of computation cost.

IV. Conclusion

This paper proposed a novel multi-target DCNN with end-to-end networks to tackle the problem of both segmentation and classification in the medical application of skin lesion detection. The result achieved by our multi-target DCNN model with three targets demonstrates that it is superior over single model like the GoogleNet or U-Net with one target. Different branches in the multi-target architecture shared detected features among different attribute categories so that they can boost each other towards a better solution. The involvement of multi-labels in the original dataset contributes to the detection of features that can describe different attributes of skin dataset, while various techniques like data augmentation and cross validation served to avoid overfitting. Since the multi-target DCNN model can be scaled to include more targets with stability [15], the target of localizing lesion dermoscopic features from ISIC 2017 can eventually be embedded in this network as future work to handle a sequence of puzzles from end to end.

Acknowledgment

The authors would like to thank Dr. Hong Liang Tey and from National Skin Center Singapore and Colin Tan from Ministry of Health Singapore for providing clinical interpretations of this study.

References

[1] R. Siegel, J. Ma, Z. Zou, and A. Jemal, 'Cancer statistics, 2014', *CA. Cancer J. Clin.*, vol. 64, no. 1, pp. 9–29, 2014.

[2] L. L. Chen, S. W. Dusza, N. Jaimes, and A. A. Marghoob, 'Performance of the first step of the 2-step dermoscopy algorithm', *JAMA Dermatol.*, vol. 151, no. 7, pp. 715–721, 2015.

[3] H. Kittler *et al.*, 'Standardization of terminology in dermoscopy /dermatoscopy: Results of the third consensus conference of the International Society of Dermoscopy', *J. Am. Acad. Dermatol.*, vol. 74, no. 6, pp. 1093–1106, 2016.

[4] J. March, M. Hand, and D. Grossman, 'Practical application of new technologies for melanoma diagnosis: Part I. Noninvasive approaches', *J. Am. Acad. Dermatol.*, vol. 72, no. 6, pp. 929-941; quiz 941-942, Jun. 2015.

[5] N. K. Mishra and M. E. Celebi, 'An overview of melanoma detection in dermoscopy images using image processing and machine learning', *ArXiv Prepr. ArXiv160107843*, 2016.

[6] A.-R. A. Ali and T. M. Deserno, 'A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data', in *Proc. of SPIE Vol*, 2012, vol. 8318, p. 83181I–1., 2012

[7] P. Carli *et al.*, 'Pattern analysis, not simplified algorithms, is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology', *Br. J. Dermatol.*, vol. 148, no. 5, pp. 981–984, 2003.

[8] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, 'Gradient-based learning applied to document recognition', *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[10] N. Codella *et al.*, 'Deep learning ensembles for melanoma recognition in dermoscopy images', *ArXiv Prepr. ArXiv161004662*, 2016.

[11] A. Esteva *et al.*, 'Dermatologist-level classification of skin cancer with deep neural networks', *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[12] ISIC 2017 challenge of skin lesion analysis, online link: https://challenge.kitware.com/#challenge/560d7856cad3a57cfde481ba. [Accessed: 1-Feb-2017].

[13] K. Matsunaga, A. Hamada, A. Minagawa, H. Koga, 'Image Classification of Melanoma, Nevus and Seborrheic Keratosis by Deep Neural Network Ensemble', *ArXiv Prepr. ArXiv170303108*, 2017.

[14] Y. Yuan, 'Automatic skin lesion segmentation with fully convolutional-deconvolutional networks', *ArXiv Prepr. ArXiv170305165*, 2017.

[15] Z. Zeng, N. Liang, X. Yang, and S. Hoi, 'Multi-target deep neural networks: Theoretical analysis and implementation', *Neurocomputing*, vol. 273, pp. 634–642, 2018.

[16] O. Ronneberger, P. Fischer, and T. Brox, 'U-net: Convolutional networks for biomedical image segmentation', in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[17] C. Szegedy *et al.*, 'Going deeper with convolutions', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015

[18] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, *Fourth International Conference on 3D Vision (3DV), IEEE*, pp. 565-571, 2016