MelaNet: A Deep Dense Attention Network for Melanoma Detection in Dermoscopy Images

Pengyi Zhang
Beijing Institute of Technology
Beijing, China
zhangpybit@gmail.com

Yunxin Zhong
Beijing Institute of Technology
Beijing, China
bityunxinz@gmail.com

Xiaoqiong Li
Beijing Institute of Technology
Beijing, China
aeople@126.com

Abstract

Skin cancer is the most common cancer worldwide, with melanoma being the most deadly form. Dermoscopy is a skin imaging modality that has demonstrated improvement for diagnosis of skin cancer compared to unaided visual inspection. However, clinicians should receive adequate training for those improvements to be realized. The success of deep learning has been witnessed as a promising technique for computer-aided biomedical image analysis. In this paper, we develop a deep neural network, i.e., MelaNet that can detect 9 different diagnostic categories of skin lesions to tackle the tasks in the ISIC Challenge 2019. MelaNet is formulated by two 169-layer dense attention networks trained on the dermoscopy images of ISIC Challenge 2019, a publicly available dermoscopy image dataset, containing 25,331 dermoscopy images with 8 diseases. We design novel dense attention block to build the dense attention networks. Such dense attention network is expected to learn more informative and discriminative deep features by fusing local and global information of images. Besides, we extend MelaNet to detect an additional outlier disease not represented in the training data through fusing multi-class classification and multi-label classification, We build an ensemble model of 4 MelaNets to further improve melanoma detection performance. These models are evaluated on validation dataset using a normalized multi-class accuracy metric. We find that MelaNet achieves state-of-the-art performance and outperforms human experts on the target task. Results of the ensemble model on ISIC2019 test dataset are eventually submitted to ISIC2019 Challenge.

Keywords: dermoscopy, melanoma detection, dense attention, multi-class classification, multi-label classification

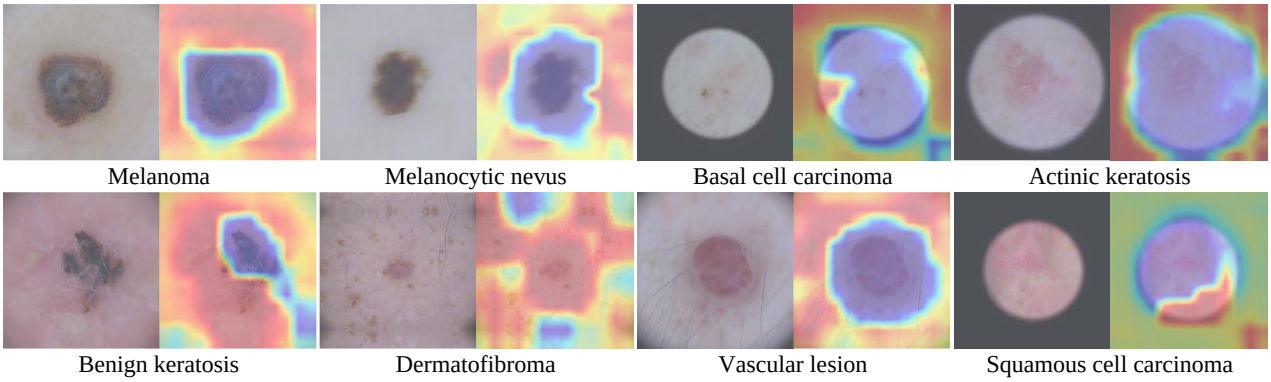

Figure 1. Instances of eight different diagnostic categories of dermatosis and their corresponding visual explanation provided by CAM [1] for the decision-making process of proposed MelaNet. The visual explanation highlights the abnormal region in dermoscopy images, presenting a high clinical relevance.

1. Introduction

Skin cancer is a major public health problem and the incidence of skin cancer continues to increase [2], thus threatening the lives of human beings worldwide. Melanoma is the deadliest form of skin cancer, responsible for an overwhelming majority of skin cancer deaths. Early diagnosis in clinical practice is able to notably improve melanoma survival. Accurate diagnosis of early melanoma particularly demands experience in dermatoscopy [3], a non-invasive examination technique that eliminates the surface reflection of skin (shown in **Figure 1**). Prior research has shown that when used by expert dermatologists, dermoscopy provides improved diagnostic accuracy, in comparison to standard photography. However, training a qualified dermatologist requires lots of money and years of efforts, which exacerbates the shortage of healthcare resources to a certain extent especially in under-developed regions. An alternative solution to improve this situation is computer-aided skin lesion analysis based on dermoscopy images.

Computer-aided biomedical image analysis has attracted many research interests. From the 1970s to the 1990s, most researchers used low-level image processing methods (e.g., edge detection, line detection, and region growth etc.) and simple mathematical models (e.g., line fitting, circle fitting, and ellipse fitting etc.) to build hybrid rule-based systems for specific biomedical image analysis tasks [4]. From the late 1990s to 2012, the methods based on traditional machine learning and pattern recognition were gradually introduced to automated biomedical image analysis and became the mainstream approaches. These methods generally consist of two core components: hand-crafted image features and pattern classifiers. The hand-crafted features mainly include local binary pattern (LBP), histogram of oriented gradient (HOG), and scale invariant feature transform (SIFT), while the frequently-used pattern classifiers include k-nearest neighbors (KNN), neural network (NN), naive Bayes classifier, and support vector machine (SVM). However, more and more complicated, fine-grained and high-level biomedical image analysis tasks have been introduced with the rapid growth of biomedical images in both resolution and modality. These approaches based on traditional machine learning and pattern recognition are critically challenged.

Driven by the growth of computing power (e.g., Graphical Processing Units and dedicated deep learning chips) and the availability of large-scale labelled samples (e.g., ImageNet [5] and COCO [6]), deep neural network has been extensively studied due to its fast, scalable and end-to-end learning framework. In recent years, Convolution Neural Network (CNN) [7] models have achieved significant improvements compared with traditional shallow methods in image classification (e.g., ResNet [8] and DenseNet [9]), object detection (e.g., Faster R-CNN [10] and SSD [11]) and semantic segmentation (e.g., UNet [12] and Mask R-CNN [13]) etc. As the biomedical imaging equipment and electronic health records (EHRs) are increasingly applied and popularized in clinical practice, a large number of biomedical images have been accumulated recently. Therefore, deep learning techniques have been gradually applied in biomedical image analysis, such as diabetic retinopathy detection [14], skin cancer diagnosis [15], lung disease detection based on X-ray radiography [16] and heart disease risk prediction on fundus retinal images [17], etc. In this paper, we aims to tackle the tasks of “Skin Lesion Analysis Towards Melanoma Detection” in the ISIC Challenge 2019 [18] using deep learning techniques. Our model, MelaNet (shown in **Figure 2**), is formulated by a multi-class classifier and a multi-label classifier, where these two classifiers are two 169-layer convolutional neural networks. MelaNet takes a dermoscopy image as input, and outputs the probability of a pathology. We train MelaNet on the dermoscopy images of ISIC Challenge 2019, a publicly available dermoscopy image dataset, containing 25,331 dermoscopy images with 8 diseases [19][20][21], including melanoma. We design novel dense attention block and use it to build dense attention network by referring to the network architecture of DenseNet [9]. Such dense attention block is expected to motivate classifiers to learn more informative deep features by fusing images’ local and global information and thus improve classification performance.Through fusing multi-class classification and multi-label classification, we extend MelaNet to detect an additional outlier disease not represented in the training data. We build an ensemble model of 4 MelaNets to further improve melanoma detection performance. These models are evaluated on validation dataset using a normalized multi-class accuracy metric. Results of the ensemble model on ISIC2019 test dataset are eventually submitted to ISIC2019 Challenge. We build an ensemble model of 4 MelaNets to further improve melanoma detection performance. These models are evaluated on validation dataset using a normalized multi-class accuracy metric. We find that MelaNet achieves state-of-the-art performance and outperforms human experts on the target task. Results of the ensemble model on ISIC2019 test dataset are eventually submitted to ISIC2019 Challenge.

2. MelaNet

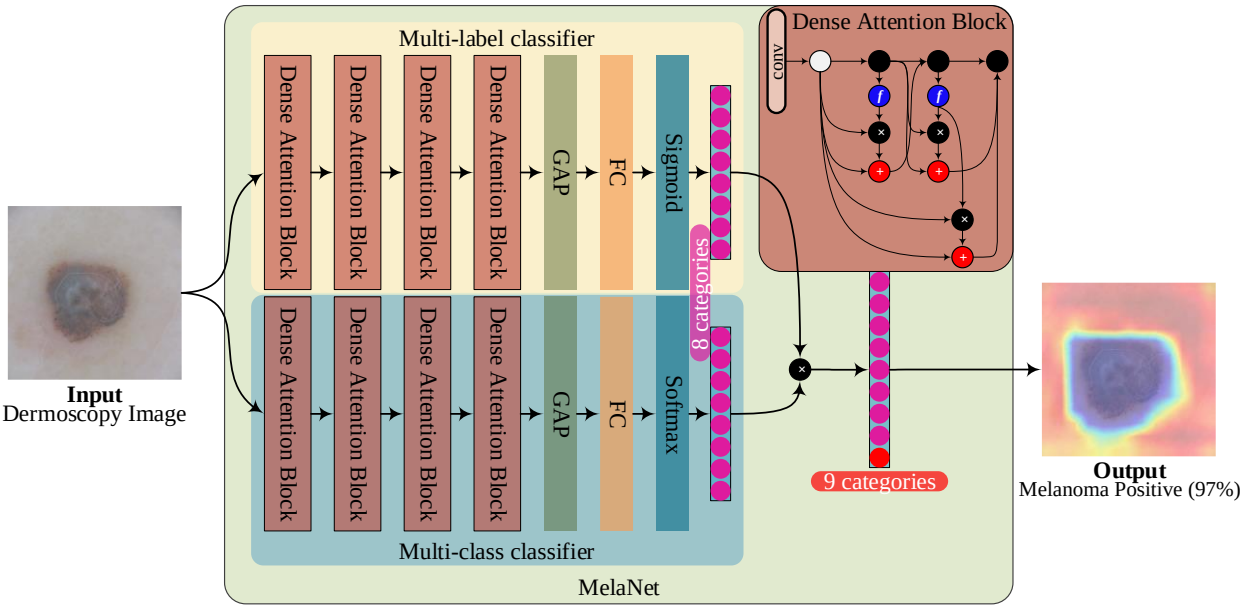

Figure 2. MelaNet is formulated by a multi-class classifier and a multi-label classifier, where these two classifiers are two 169-layer dense attention networks trained on the dermoscopy images of ISIC Challenge 2019. MelaNet takes a dermoscopy image as input, and outputs the probability of a pathology. On this example, MelaNet correctly detects melanoma and also localizes areas in the image most indicative of the pathology.

2.1 Problem Formulation

In the task of "Skin Lesion Analysis Towards Melanoma Detection" in ISIC Challenge 2019, developed algorithms require to identify 9 different diagnostic categories of skin lesions, where only 8 categories are represented in the training data. First, we formulate this task as a multi-class classification problem, where the input is a dermoscopy image X and the output is an octonary label $y \in \{0,1,...,7\}$ indicating which category of skin lesions. For a single example in the training set, we optimize the weighted cross entropy loss

$$L(X, y) = \sum_{c=0}^{7} -w[c] \cdot y_{one-hot}[c] \log p_s(Y = c \mid X), \quad (1)$$

where $p_s(Y = c \mid X)$ is the probability that the network assigns to the label c, $y_{one-hot}[c]$ is the one-hot encoding of y and $w[c] = 1 - N_c / N$ with N_c and N the number of dermoscopy images with the label c and total number of dermoscopy images in the training set respectively. Second, we formulate this task as a multi-label classification problem to support identification of an additional outlier class not represented in the training data, where the input is a dermoscopy image X and the output is an octonary label $y \in \{0,1\}$ indicating the absence or presence of a specific category of skin lesions. For a single example in the training set, we optimize the weighted binary cross entropy loss

$$L(X, y) = -w_+ \cdot y \log p_b(Y = c \mid X) - w_- \cdot (1 - y)(1 - \log p_b(Y = c \mid X)), \quad (2)$$

where $p_b(Y = c \mid X)$ is the probability that the network assigned to the label c, $w_+ = N_- / (N_+ + N_-)$ and $w_- = N_+ / (N_+ + N_-)$ with N_+ and N_- the number of dermoscopy images with and without the label c in the training set respectively. Finally, we fuse the results of multi-class classification and multi-label classification, i.e., $p_s(Y = c \mid X)$ and $p_b(Y = c \mid X)$ to formulate the classification results of 9 different diagnostic categories of skin lesions

$$p(Y=c \mid X)=\begin{cases}1\text{-}(1\text{-}p_s(Y=c \mid X))\cdot(1-p_b(Y=c \mid X)),\ c=0,1,...,7\\ \prod_{c=0}^{7}(1-p_s(Y=c \mid X)\cdot p_b(Y=c \mid X)),\ c=8\end{cases}, \tag{3}$$

2.2 Model

Dense attention network. We design novel dense attention block and use it to build dense attention networks by referring to the architecture of DenseNet [9]. Our dense attention block contains an additional attention branch in the architecture compared to dense connections in DenseNet (formula 4). We formulate the attention branch as a residual term (formula 5).

$$x_l = H_l([x_0,...,x_{l-2},x_{l-1}]) \ , \tag{4}$$

$$\begin{aligned} a_{l\cdot k} &= f(x_{l-k-1})\cdot x_k \\ x_l &= H_l([x_0+a_{l\cdot 0},...,x_{l-2}+a_{l\cdot(l-2)},x_{l-1}]) \end{aligned}, \tag{5}$$

where x_l is the deep feature of l-th layer in dense attention block, $a_{l\cdot k}$ is the attention residual of l-th layer and k-th layer and f denotes normalization function. We empirically take softmax function in spatial space of feature maps as f in our implementation. We argue that high layers in dense attention network extract semantic and global information. Such semantic and global information is used as local information guidance of low layers. Multiplication operation between features from high layers and low layers can enhance semantically related local information, called attention in this paper. Specifically, we use a convolutional layer with 1×1 filter kernel to maintain the consistency in the channel number of feature maps in multiplication operation.

MelaNet. MelaNet is formulated by a multi-class classifier and a multi-label classifier as shown in **Figure 2**. We take two 169-layer dense attention networks as backbones. We then use fully connected (FC) layer that have eight output units to build the multi-class classifier and multi-label classifier, after which we apply a "softmax" nonlinearity and a sigmoid nonlinearity respectively. We obtain the output of MelaNet by fusing the outputs of these two classifiers according to **formula 3**. In implementation of MelaNet, we actually use eight individual binary classifiers to realize the multi-label classifier instead of using the common multi-task framework that consists of one shared backbone network and eight individual classifier headers. We argue that in spite of consuming more computing resources and running slower, training eight binary classifiers separately is more competitive than training the common multi-task framework. In the training process of each individual binary classifier, the optimizer only needs to work on one single binary classification problem regardless of the optimization of other categories. Such exclusive optimization mechanism is expected to be complementary with the joint optimization of multi-class classifier in the formulation of MelaNet. Besides, in order to make training process converge fast and stably, we initialize the weights of these eight binary classifiers with the weights of trained multi-class classifier. Finally, we build an ensemble model of 4 MelaNets to further improve melanoma detection performance.

3. Experiments

MelaNet is implemented based on the DenseNet implementation in torchvision package[1]. We use a Linux server with Intel(R) Xeon(R) E5-2683 v3 CPU @ 2.00GHz (56 CPUs), 64GB RAM, and four NVIDIA GTX1080ti GPU cards to train and evaluate models in our experiments.

3.1 Dataset

The available data in ISIC Challenge 2019 contains 25,331 dermoscopy images with 8 diseases [19][20][21], including melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, vascular lesion, and squamous cell carcinoma. In our experiments, we randomly split the data into training (80%, 20269 images) and

[1] https://github.com/pytorch/vision/tree/master/torchvision

validation (20%, 5062 images), where the split ratio (4:1) is across all 8 categories. All models in this paper are trained on training set and evaluated on validation set.

3.2 Training

Multi-class classifier. We first train the multi-class classifier on the training set. The weights of the network are initialized with weights pretrained on ImageNet. The network is trained end-to-end using Adam with standard parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$). We train the model with 20 epoches using mini-batches of 8. We use an initial learning rate of 0.001 that is decayed by a factor of 10 each time the validation lass plateaus after an epoch, and pick the model with lowest validation loss.

Multi-label classifier. We then train eight individual binary classifiers on the training set separately. The weights of these networks are initialized with weights of trained multi-class classifier. These networks are trained end-to-end using Adam with standard parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$). We train these models with 10 epoches using mini-batches of 8. We use an initial learning rate of 0.0001 that is decayed by a factor of 10 each time the validation lass plateaus after an epoch, and pick the model with lowest validation loss.

Before inputting the images into these networks, we resize the images to 416×416 and normalize based on the mean and standard deviation in the ImageNet training set. We also augment the training data with random horizontal flipping and random rotation between -30 and 30 degrees.

3.3 Test

We test these models on validation set with a normalized multi-class accuracy metric, sensitivity, specificity, negative predictive value and positive predictive value. Normalized (or balanced) multi-class accuracy is defined as the accuracies of each category, weighted by the category prevalence. We evaluate these models using the publicly available "isic-challenge-scoring" package[2] provided by ISIC Challenge 2019.

Finally, results of the ensemble MelaNet on ISIC2019 test dataset are submitted to ISIC2019 Challenge.

4. Results

We report the diagnostic performance of proposed MelaNet per diagnostic category in **Table 1**. We also produce heatmaps to visualize the areas of the image most indicative of the disease using class activation mappings (CAMs) as shown in **Figure 2**. The visual explanation highlights the abnormal region in dermoscopy images, presenting a high clinical relevance.

Table 1. Diagnostic performance of proposed MelaNet per diagnostic category.

Pathology	Sensitivity	Specificity	Negative predictive value	Positive predictive value	Normalized accuracy
Melanoma	79.50%	96.30%	95.60%	82.30%	85.90%
Melanocytic nevus	94.70%	91.60%	94.40%	92.10%	
Basal cell carcinoma	94.10%	98.10%	99.10%	88.30%	
Actinic keratosis	83.80%	98.10%	99.40%	60.90%	
Benign keratosis	83.80%	97.50%	98.10%	79.20%	
Dermatofibroma	85.10%	99.40%	99.90%	58.80%	
Vascular lesion	96.00%	99.80%	100.00%	82.80%	
Squamous cell carcinoma	80.00%	99.10%	99.50%	70.40%	

5. Conclusion

The incidence of skin cancer continues to increase worldwide with melanoma being the most deadly form. Accurate diagnosis of early melanoma particularly demands experience in dermatoscopy. However, clinicians should receive adequate training for those improvements to be realized. In this paper, we develop a deep neural network, i.e., MelaNet for melanoma detection in dermoscopy images. MelaNet is formulated by two 169-layer convolutional neural networks

[2] https://github.com/ImageMarkup/isic-challenge-scoring/

trained on the dermoscopy images of ISIC Challenge 2019, a publicly available dermoscopy image dataset, containing 25,331 dermoscopy images with 8 diseases. Through fusing multi-class classification and multi-label classification, we extend MelaNet to detect an additional outlier disease not represented in the training data. We build an ensemble model of 4 MelaNets to further improve melanoma detection performance. These models are evaluated on validation dataset using a normalized multi-class accuracy metric. We find that MelaNet achieves state-of-the-art performance and outperforms human experts on the target task. Results of the ensemble model on ISIC2019 test dataset are eventually submitted to ISIC2019 Challenge.

References

[1]. Zhou, Bolei, et al. Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 2921-2929.

[2]. Mishra, Nabin K., and M. Emre Celebi. "An overview of melanoma detection in dermoscopy images using image processing and machine learning." arXiv preprint arXiv:1601.07843 (2016).

[3]. Tschandl, Philipp, et al. "Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study." Lancet Oncology (2019).

[4]. Litjens G , Kooi T , Bejnordi B E , et al. A survey on deep learning in medical image analysis[J]. Medical Image Analysis, 2017, 42:60-88.

[5]. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 248–255. 1

[6]. Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//European conference on computer vision. Springer, Cham, 2014: 740-755.

[7]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol.521, no. 7553, pp. 436–444, 2015. 1

[8]. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[9]. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[10]. Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.

[11]. Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[12]. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[13]. He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

[14]. Gulshan V , Peng L , Coram M , et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs[J]. JAMA, 2016.

[15]. Esteva A , Kuprel B , Novoa R A , et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2017, 542(7639):115-118.

[16]. Rajpurkar P, Irvin J, Ball R L, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists[J]. PLoS medicine, 2018, 15(11): e1002686.

[17]. Poplin R , Varadarajan A V , Blumer K , et al. Predicting Cardiovascular Risk Factors from Retinal Fundus Photographs using Deep Learning[J]. Nature Biomedical Engineering, 2017.

[18]. Codella, Noel C. F., Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen W. Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael A. Marchetti, Harald Kittler and Allan Halpern. "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)." ArXiv abs/1902.03368 (2019): n. pag.

[19]. Tschandl P., Rosendahl C. & Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi.10.1038/sdata.2018.161 (2018)

[20]. Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi

Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern: “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)”, 2017; arXiv:1710.05006.

[21]. [3] Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Allan C. Halpern, Susana Puig, Josep Malvehy: “BCN20000: Dermoscopic Lesions in the Wild”, 2019; arXiv:1908.02288.