DEEPMOLE: DEEP NEURAL NETWORKS FOR SKIN MOLE LESION CLASSIFICATION

V. Pomponiu, H. Nejati, N.-M. Cheung

Information System, Technology and Desing (SUTD)
Singapore University of Technology and Design (ISTD)
Somapah Road 8, Singapore, 487372

ABSTRACT

Nowadays, the occurrence of skin cancer cases has grown worldwide due to the extended exposure to the harmful radiation from the Sun. Most common approach to detect the malignancy of skin moles is by visual inspection performed by an expert dermatologist, using a set of specific clinical rules. Computer-aided diagnosis, based on skin mole imaging, is another concurrent method which has experienced major advancements due to improvement of imaging sensors and processing power. However, these schemes use hand-crafted features which are difficult to tune and perform poorly on new cases due to lack of generalization power. In this study we present a method that use a pretrained deep neural network (DNN) to automatically extract a set of representative features that can be later used to diagnose a sample of skin lesion for malignancy. The experimental tests carried out on a clinical dataset show that the classification performance using DNN-based features performs better than the state-of-the-art techniques.

Index Terms— Skin mole classification, malignant melanoma, deep neural networks, transfer learning, feature extraction.

1. INTRODUCTION

Malignant melanoma (MM) [1] is a type of skin cancer characterized by an excessive development of the epidermis cells. If diagnosed in an early stage is curable; thus, early detection of melanoma is the key since the 5 years survival rates drop from 98% for localized melanoma to 15% for the metastatic disease [2, 3].

In the last 30 years the incidence of MM has been growing rapidly and steadily, and therefore there is an urgent need for tools that can monitor and track suspicious skin lesions. The process of differentiating melanoma from benign nevi is complex and involves combining patient-derived anapestic data, analytical reasoning, comparative and differential recognition, and pattern analysis [4]. Examples of benign nevi and MM images are shown in Fig. 1.

Throughout time, several clinical rules devised by dermatologists have been established in order to classify the skin lesions. The three common rules are the ABCD rule [1], the 7-point list [5], and the Menzies rule [6].

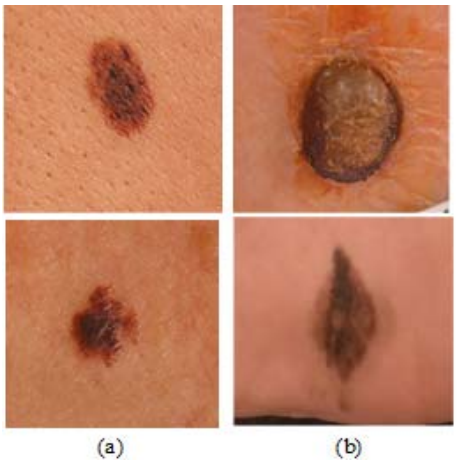

Fig. 1. Illustration of different skin lesions: (a) benign nevi, (b) MMs.

The ABCD rule [1] consists of four features: **A**symmetry of lesion, **B**order irregularity, **C**olor variation, **D**iameter and **E**volving. However, the ABCD rule has some limitations: is not able to distinguish some dysplastic nevi from MMs and to identify some MMs at a very early stage [7]. In later stage, the skin lesion texture was emphasized to be sensitive to MM identification [8, 9].

The 7-point check list [5] is based on seven dermatoscopic features such as atypical pigment network, blue-whitish veil, atypical vascular pattern, irregular streaks, irregular pigmentation, irregular dots/globules and regression structures. The first three features are called major criteria while the other four are minor criteria. Instead, the Menzies rule [6] relies on the definitions of negative features (e.g., uniform color pattern distribution) and positive features (e.g., multiple brown dots, multiple colors and broadened network).

Computer-aided schemes for melanoma classification can be constructed by adopting a two step process [6]: first, computing a compact set of discriminative features which describe the mole region and then building a classification model for the skin lesions based on the extracted features.

There is a plethora of computer-aided systems for classification of skin mole lesions. The features used to accurately classify MM from dermatoscopic images are devised in such a way that they can describe dermatologist-observed characteristics [7]. There are many methods used in the literature that capture these features [10, 11]. For instance, in [12] an artificial neural network (ANN) classifier is fed with a set of discriminant features, based on tumor shape and relative tumor color. Saez et al. [13] proposed a model-based classification of the global dermatoscopic patterns. The method employs a finite symmetric conditional Markov model in the color space and the resulted parameters are treated as features. In [14] a layered classification model, using three feature categories (color, subregion and texture), is used for a wide range of skin lesions, including non-melanocytic ones.

Recently, several methods extended the analysis from dermatoscopic images to high-resolution images of skin lesions captured by a standard camera or smartphone [15–19]. For instance, Wadhawan et al. [15] proposed a portable library for melanoma detection on handheld devices based on the well-known bag-of-features framework. Doukas et al. [17] also focuses on camera images. The lesion detection and feature extraction are performed on mobile device while the classification can be performed on the cloud. In [18] several features are proposed which are able to capture the ABCDE rule, while in [19] local patterns are used for texture characterization. Techniques which employ neural networks have appeared recently [20]. However due to data limitations in histology slides these approaches build a small-scale neural network with very few layers.

Although DNN can be used to perform classification directly using the output of the last network layer, they can also behave as feature extractors combined with a classifier. In this paper, we investigate a classification system that employs a DNN, pretrained using natural images, to extract features suitable to another domain, i.e., dermatology, where there are few annotated samples available. The gain in classification accuracy is shown compared to methods that use hand-crafted features.

The rest of the paper is organized as follows. In Section 2 we give a detailed explanation of DNN-derived features generated automatically from a dataset of skin mole images. Furthermore, the section describes the DNN architecture, the techniques used to reduce overfitting and the employed classifier. The performance evaluation of the proposed classification system is reported in Section 3, while Section 4 presents the conclusion and the feature work.

2. THE PROPOSED METHOD

At the core of our system is a pretrained convolutional neural network (CNN), which is used as a feature extractor for our skin cancer dataset. This process, which transfer the knowledge of an existing CNN to a new domain, have been recently studied and proved successful for several applications [21, 22]. Our main motivation of using this approach instead of fine-tuning the CNN [22] is determined by several factors such as the size and the type of our dataset. Therefore, since our dataset is rather small the best option is to build a classier model on top of the output (activations) of the hidden layers. Furthermore due to the mismatch between the dataset that was used to train the CNN and our dataset, is better to construct the classifier using the activations belonging to earlier layers of the network. The block diagram of the proposed detection system based on DNN is summarized in Fig. 2.

2.1. The CNN Architecture

The architecture of the CNN used as a feature extractor is the one proposed in [23]. As illustrated in Fig. 2 (i.e., the dashed red block), the net comprises of eight layers; the first five are convolutional layers and the remaining three are fully connected layers.

The first convolutional layer filters the input color image with 96 kernels of size $11 \times 11 \times 3$. The second convolutional layer takes as input the output (normalized and pooled) of the first convolutional layer and filters it with 256 kernels of size $5 \times 5 \times 48$. The third, fourth, and fifth convolutional layers are chained to one another without applying any pooling or normalization layers. Instead, the sixth and seventh fully-connected layers have each 4096 activation neurons. For a complete description of the features of the CNN architecture, the reader is invited to consult [23].

In this paper we consider the general image classification scenario: given a skin mole color image, $I^{N\times M\times 3} \in L^2(\Re^d)$ (where $L^2(\Re^d)$ stands for the space of Lebesgue-measurable functions), predict if it is benign or melanoma, with the additional constraint that the classification output should match the groundtruth diagnosis (i.e., the histological result).

Following the mathematical construction given in [24], the overall features extracted by the CNN are given by

$$F(I) = \bigcup_{n=0}^{\infty} F^n(I) \tag{1}$$

an those corresponding to the n-th layer are

$$F^n(I) = U_n[q]I \star \chi_n \tag{2}$$

where q is a path, i.e., an ordered sequence defined on the countable index set $q = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \Lambda_1^n$, and $\Lambda_1^n = \Lambda_1 \times \Lambda_2 \times \ldots \times \Lambda_n$.

The operator $U_n = ((M_n(I \star g_n))(R_n)$ encompasses several operations when applied on I, such as convolution with the atoms g_n take from a general semi-discrete shift-invariant frame, followed by a non-linearity $M_n : L^2(\Re^d) \rightarrow L^2(\Re^d)$ which satisfies the Lipschitz property:

$$||M_n I - M_n J|| < ||L_n(I - J)||_2 \tag{3}$$

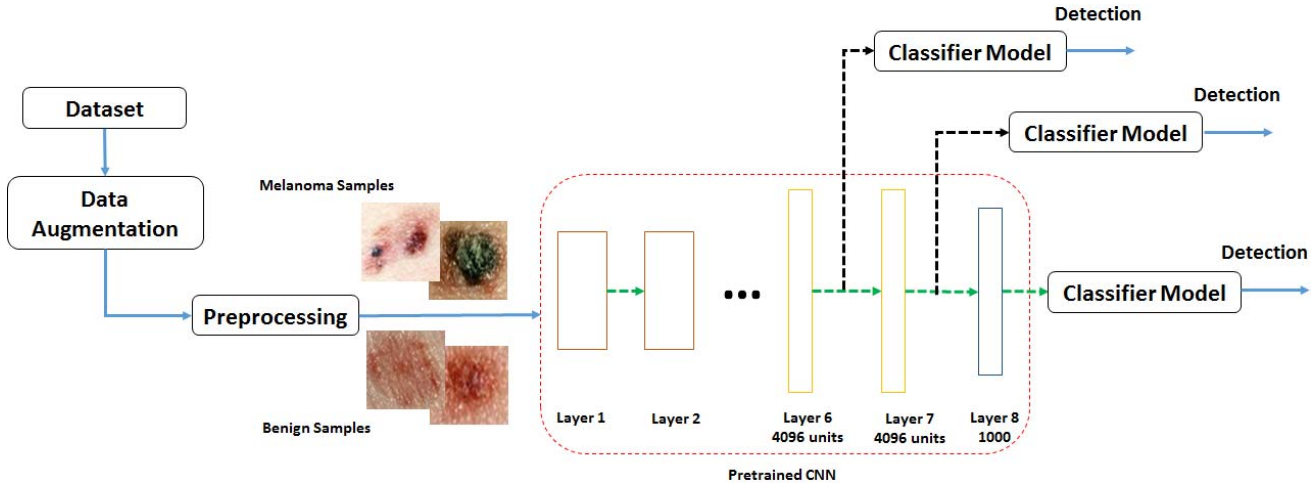

Fig. 2. An illustration of the proposed system for skin cancer detection. The "Data Augmentation" block aims to enlarge the dataset by employing a set of distortions to avoid overfitting and in the same time to achieve feature robustness. The skin mole images, belonging to two classes, are resized and normalized before presented at the input of the pretrained CNN. The features extracted from the last three layers are used by the classifier model to output a decision whether the input image is benign or malignant.

$I, J \in L^2(\Re_d)$, and with $M_n I = 0$ for $I = 0$. The output of the linearity is further subsampled by a certain factor $R_n \geq 1$, i.e., sub-sampling by a factor of R aims to retain only every R-th sample. The function χ_n is the output-generating atom of the n-th layer, where for n = 0 it represents to the root of the network.

In our system we decided to use the features extracted from the last three layers (i.e, $F^n(I)$ with $n = \{6, 7, 8\}$ of the CNN and build a custom classifier for each of these feature sets. It is worth to point out that, for all the images in the dataset the region of interest (ROI) which contains the skin mole has been already outlined. In order to use the CNN with our dataset we need to match the net input; thus, from each ROI image in the dataset we first subtract the mean image [23] learned by the net, and then resize it to 224 $\times$ 224.

2.2. Data Augmentation to Reduce Overfitting

Since the CNN architecture has a very large number of parameters there is a significant risk of overfitting. In order to reduce the overfitting effect we artificially expand our dataset of digital camera images of skin moles.

We employ a set of distortions that modifies the intensities of RGB color bands of the images. The employed distortions are Gaussian low pass filter, histogram equalization, noise addition, guided filtering, JPEG compression and motion blur.

Another motivation of augmenting the dataset by including these distortions is to gain feature invariance by generating image samples that can resemble the distorted image captured by smartphone in loosely acquired conditions (e.g., blurring due to device shaking or changes in the intensity due to illumination).

2.3. The Classifier

The classifier model that we choose to build on the top of learned features extracted from the last three layer of the CNN is the k nearest neighbor classifier (kNN) since it is a good candidate when working in a distance representation of objects.

We adopt kNN with the distance metric between two images being a pooled vector difference between extracted features. Each ROI image (segmented skin mole) is denoted by a feature vector. Dissimilarities between ROIs are expressed as dissimilarities between the feature vectors:

$$D(I, J) = \sum_{p=1}^{m} D(F_p^n(I), F_p^n(J)) \tag{4}$$

where are I and J are the ROIs of two skin mole images, F_p^n denotes the corresponding feature vectors and D is the similarity metric.

In general, several metrics have been employed to classify the extracted features such as, chi-square distance, L1 distance and cosine distance. In this study we decided to adopt the cosine distance metric for the kNN classifier since it is more robust to outliers and it was widely used in many previous works.

3. SIMULATION RESULTS

3.1. The Dataset

The dataset used in this paper has been collected from DermIS [25] which is an European dermatology atlas for

Table 1. The average performance across all folds and the associated standard deviation of the proposed system based on DL-features, in terms of sensitivity, specificity and accuracy, compared with the methods based on hand-crafted features.

Method	[16]	Doukas et al. [17]	Do et al. [18]	Pomponiu et al. [19]	F^6	F^7	F^8
Sens (%)	60.7±5.5	64.11±4.6	84.83±2.4	85.71±2.1	88.46±0.8	87.91±1.7	92.10±0.5
Spec (%)	80.5±3.8	81.67±7.1	86.32±8.4	82.48±6.8	93.08±1.1	94.93±1.0	95.18±1.3
Acc (%)	66.7±0.7	77.06±2.9	85.58±3.9	83.95±3.3	90.77±1.5	91.42±0.6	93.64±1.9

healthcare professionals, and another popular service, DermQuest [26], which provides an online dermatology image library.

Our dataset consists of 399 color (RGB) images of skin mole lesions acquired using a digital camera with different resolutions. The image dataset is classified into two classes: 217 benign nevi and 182 MMs. After applying the data augmentation we enlarged the dataset to 10k images.

The diagnosis of the melanoma cases were obtained from histology through the biopsy of the skin lesions. In order to obtain the ROI for each skin lesion we manually annotate (label) them. In the next section we present the classification results obtained for our dataset while using DL-based features and hand-crafted one.

3.2. Classification results

In order to assess the performance of the proposed system, we choose hand-crafted feature sets, F, that are employed by several recent skin cancer detection methods [16–19]. We implemented them by following the detailed description give in the papers were their were introduced.

The features proposed in [16–19] were extracted from ROI skin images without any preprocessing (such as noise removal). Furthermore, we resized the ROIs to 256×256 pixels using cubic interpolation with anti-aliasing enabled.

To standardize the range of the computed features we normalize them using the z-score: z-score $= (F - \mu)/\sigma$, where μ and σ represent the mean and standard deviation of feature vector F for the entire dataset.

The parameters used for the kNN classifier, determined after a grid search, were the number of neighbors $k = 2$ and the distance between the feature vectors $D = cosine$. The performance of the classifier is calculated in terms of sensitivity (i.e., Sens = TP/(TP + FN)), specificity (i.e., Spec = FN/(FN + FP)) and accuracy (i.e., Acc = (TP + TN)/(TP + FP + TN + FN)).

To estimate the generalization error of the classifiers models, that use the learned feature set, we employed the 10-fold cross validation. In each fold trial all the samples of the 1 fold are held out and used for testing while the remaining samples, belonging to the remaining folds, are used for training.

The implementation of the CNN architecture used for transfer learning is done with the aid of Caffe deep learning library [27]. We use the net introduced in [23] together with corresponding mean image learned by the model during pretraining.

Table 1 reports the results of the proposed method and four other recent state-of-the-art methods based on hand-crafted features. The columns denoted by F^n, represent the features extracted from the last three layers of the pretrained CNN.

It can be observed that the proposed method shows better performance compared to the methods that employ hand-tunned features. Furthermore, is interesting to note that even if the CNN has been tarined on a different dataset (images of common objects), which is far from the skin cancer domain, the features transfered from the last layers are able preserve the discrimination between the classes. The main difference between these domains is the content which can change the skin tones due to the lighting conditions.

4. CONCLUSION

In this paper we investigate the transfer learning capability of the features extracted from a pretrained CNN for the classification of skin lesion images.

The experimental results show that the transfered features are a viable solution for skin mole characterization, achieving better classification accuracy than method that make use of had-crafted features.

Although the net was trained using the images coming form another domain, it is possible to transfer the learning to another different domain, i.e., skin cancer, which opens an interesting venue for skin cancer research.

In the future we plan to investigate two research directions: the first aims to combine the DNN-based features obtained from different layers of the CNN or with the features that describe the well-know dermatological rules. We believe, this venue of research will improve the classification performance fo the skin cancer detection system.

The second direction explores the adaption of these features to other type of skin cancer and even the general case of skin assessment done through mobile devices which will enable easy triage of the patients. For these power-constrained mobile devices the feature computation should be lightweight and fast, requirements that can be achieved through feature compression and GPU-based acceleration.

5. REFERENCES

[1] American Academy of Dermatology, "Melanoma: Signs and symptoms," https://www.aad.org/dermatology-a-to-z/diseases-and-treatments/m---p/moles/signs-symptoms, Accessed on Feb 1, 2016.

[2] American Cancer Society, "Cancer facts & figures," http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-044552.pdf, Accessed on Feb 1, 2016.

[3] A. Börve et al., "Smartphone teledermoscopy referrals: A novel process for improved triage of skin cancer patients," *Acta Derm Venerol*, vol. 95, pp. 186–190, 2015.

[4] A.A. Marghoob and A. Scope, "The complexity of diagnosing melanoma," *J Invest Dermatol*, vol. 129, pp. 11–13, 2009.

[5] I. Maglogiannis and C.N. Doukas, "Overview of advanced computer vision systems for skin lesions characterization," *IEEE TITB*, vol. 13, pp. 721733, Sept. 2009.

[6] S.W. Menzies, C. Ingvar, and W.H. McCarthy, "A sensitivity and specificity analysis of the surface microscopy features of invasive melanoma," *Melanoma Research*, vol. 6, pp. 55–62, 1996.

[7] R. Amelard et al., "High-level intuitive features (hlifs) for intuitive skin lesion description," *IEEE TBME*, vol. 62, pp. 820–831, Mar. 2015.

[8] C. Barata et al., "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE J SYST*, vol. 8, pp. 965–979, 2014.

[9] G. Argenziano et al., "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions," *Archives of Dermatology*, vol. 134, Dec. 1998.

[10] P. Sabouri, H. Gholam, T. Larsson, and J. Collins, "A cascade classifier for diagnosis of melanoma in clinical images," *36th IEEE EMBS*, Aug. 2014.

[11] S. Kia et al., "Computer-aided diagnosis (cad) of the skin disease based on an intelligent classification of sonogram using neural network," *Neural Computing and Applications*, vol. 22, no. 6, pp. 1049–1062, 2013.

[12] F. Ercal et al., "Neural network diagnosis of malignant melanoma from color images," *IEEE TBME*, vol. 41, pp. 837–845, 1994.

[13] A. Saez, C. Serrano, and B. Acha, "Model-based classification methods of global patterns in dermoscopic images," *IEEE TMI*, vol. 33, pp. 1137–1147, May 2014.

[14] K. Shimizu et al., "Four-class classification of skin lesions with task decomposition strategy," *IEEE TBME*, vol. 62, pp. 274–283, Jan. 2015.

[15] T. Wadhawan et al., "Implementation of the 7-point checklist for melanoma detection on smart handheld devices," *IEEE EMBS*, Aug. 2011.

[16] Ramlakhan K. and Y. Shang, "A mobile automated skin lesion classification system," in *ICTAI*, 2011, pp. 138–141.

[17] Doukas Charalampos et al., "Automated skin lesion assessment using mobile technologies and cloud platforms," in *Conf Proc IEEE EMBS*, 2012, pp. 2444–2447.

[18] T.T. Do et al., "Early melanoma diagnosis with mobile imaging," in *IEEE EMBC*, 2014, pp. 6752–6757.

[19] V. Pomponiu, H. Nejati, and N.-M. Cheung, "Texture analysis of skin mole lesions using discriminative LBP," in *24th EADV Congress*, 2015, pp. 1–6.

[20] A.C Roa et al., "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," in *SPIE Medical Imaging*, 2014, vol. 9041, pp. 1–15.

[21] P. Lamblin and Y. Bengio, "Important gains from supervised fine-tuning of deep architectures on large labeled sets," NIPS, 2010.

[22] J. Yosinski et al., "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani et al., Eds., pp. 3320–3328. Curran Associates, Inc., 2014.

[23] A. Krizhevsky et al., "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira et al., Eds., pp. 1097–1105. Curran Associates, Inc., 2012.

[24] T. Wiatowski and B. Bölcskei, "Deep convolutional neural networks based on semi-discrete frames," *CoRR*, vol. abs/1504.05487, 2015.

[25] DermIS, "Online dermatology internet service," http://www.dermis.net, Accessed on January 31, 2016.

[26] DermQuest, "Online medical resource," http://www.dermquest.com, Accessed on January 31, 2016.

[27] Y. Jia et al., "Caffe: Convolutional architecture for fast feature embedding," in *22nd ACM MM*, 2014, MM '14, pp. 675–678.