Simple and Effective Pre-processing for Automated Melanoma Discrimination based on Cytological Findings

Takuya Yoshida*, M. Emre Celebi†, Gerald Schaefer‡ and Hitoshi Iyatomi*
*Department of Applied Informatics, Graduate School of Science and Engineering, Hosei University, Japan,
e-mail: takuya.yoshida.3d@stu.hosei.ac.jp, iyatomi@hosei.ac.jp
†Department of Computer Science, University of Central Arkansas, e-mail: ecelebi@uca.edu
‡Department of Computer Science, Loughborough University, U.K., e-mail: gerald.schaefer@ieee.org

***Abstract*—In this paper, we propose a simple and effective pre-processing method for melanoma classification by considering cytological properties of melanomas, in particular the alignment of the major axis of the tumor in the same direction. We evaluate our method with a set of 1,760 dermoscopic images (329 of melanomas and 1,431 of nevi) and a simple convolutional neural network (CNN) classifier with five-fold cross validation. The proposed tumor alignment method improves the classification performance by 5.8% in terms of the area under the ROC curve (AUC). In addition, it proves to be 2.1% better in term of AUC when compared with the same configured CNN trained using images that are nine times larger. Our results also show that considering the intrinsic features of the classification target is important even when the classifier has a capability to obtain effective features automatically through its learning process.**

I. Introduction

Over the past few decades, the incidence of malignant melanoma has gradually increased in most parts of the world. The latest statistics from Australia reveal that the incidence of melanoma is 53 cases for every 100,000 people, and that 1,617 people died from malignant melanoma in 2012 [1]. Melanoma quickly metastasizes and thus early detection is critical to reducing melanoma-related deaths. However, distinguishing between early stage melanoma and nevus (i.e., benign pigmented skin lesion) is often difficult. Dermoscopy, a non-invasive skin imaging technique, has been shown to improve the accuracy of melanoma diagnosis [2], with a systematic review covering Medline entries revealing that it has 10-27% higher sensitivity [3]. Dermoscopy has consequently become quite popular in dermatology. However, dermoscopic diagnosis requires extensive training and experience, and despite the use of dermoscopy, the accuracy of expert dermatologists in diagnosing melanoma is estimated to be only 75-88% [4], [5].

Several groups have developed automated analytical procedures to overcome these problems and have reported high rates of diagnostic accuracy [6]–[8]. We also have been developing an Internet-based melanoma screening system [6] (available at http://dermoscopy.k.hosei.ac.jp). Our system achieved a sensitivity (SE) of 85.9% coulpled with a specificity (SP) of 86.0% on a set of 1,258 non-acral dermoscopy images (1,060 melanocytic nevi and 198 melanomas) using an artificial neural network. Liu *et al.* [7] focused on the symmetricity of a tumor and developed a pigmentation elevation model and global point signatures. Their model achieved high classification accuracy (SE = 86.4% and SP = 82.1%) with only four features. Shimizu *et al.* [8] developed a classification model using a task decomposition strategy that supports not only melanocytic lesions but also non-melanocytic skin lesions, which are commonly seen in clinical practice. The model simultaneously achieved detection rates of 90.5%, 82.5%, 82.6%, and 80.6% for melanomas, nevi, basel cell carcinomas, and seborrheic keratosis, respectively.

In recent years, deep learning, a new machine learning schema, has gained many promising achievements in a wide range of applications. A convolutional neural network (CNN) [9] is a principal aspect of deep learning techniques specialized for machine learning and/or computer vision. The most attractive advantage of CNNs over conventional classification techniques is that they extract effective features for the classification task from given training images automatically as part of their learning scheme. Because CNNs automatically perform this extraction process, they considerably reduce the burden of the challenging task of designing effective features. Several groups have recently applied these methodologies to their melanoma classification systems. Codella *et al.* [10] emphasized the analogy between the skin's structure and a natural scene and pre-trained their CNN with natural images. They performed two kinds of two-class classification tasks, melanoma vs. non-melanoma and melanoma vs. atypical lesions, and achieved 93.1% and 73.9% in average accuracy, respectively. These results outperformed conventional approaches evaluated by the same dataset. Later, Kawahara *et al.* [11] also pre-trained their CNN with natural images from an ImageNet dataset and evaluated the performance. They extracted pre-trained CNN to fully convolutional layers. This was accomplished by converting the fully connected layers of the CNN to convolutional layers using techniques recently applied to the general computer vision technique [12]. They achieved a performance of 81.8% classification accuracy on a 10 class dataset. These methodologies also employ the pre-training process to improve performance.

978-1-4673-9005-7/16/$31.00 ©2016 IEEE

CNNs thus have major advantages and attain good, often excellent, performance even without considering the property of the target. However, one disadvantage of CNNs is that they require considerable training data or data augmentation to avoid overfitting. We believe it is quite important to consider the intrinsic property of the target, namely the dermatological property of pigmented skin tumors. In this paper, we therefore propose a simple and effective pre-processing method for melanoma classification while considering the dermatological nature of pigmented skin tumors and confirm the advantages of our presented approach.

II. Methods

A. Proposed pre-processing method based on cytological findings for melanoma classification

Based on medical knowledge, the shape of nevi tends to become symmetrical because cells of nevus grow at a regular speed. On the other hand, cancer cells grow and spread vigorously and randomly, and often show increased asymmetry in their appearance. This tendency is encapsulated as the (A) (asymmetry of the tumor) in the commonly used ABCD rule [13] for melanoma diagnosis, which quantifies the asymmetry (A), border sharpness (B), color variegation (C) and the number of differential structures (D) present in a lesion (see Appendix for more details). Based on this cytological finding, our quite simple idea is aligning the major axis of the tumor before the subsequent processes, i.e. feature extraction and classification. We think this alignment helps to produce intrinsic features of the tumor to be classified.

Fig. 1 shows a schematic of the proposed process. Our algorithm is based on a simplified version of our tumor extraction method from [14]. It consists of three phases: **1) Thresholding** where we binarize the input image using the traditional Otsu thresholding method on the blue channel, **2) Regionalization** where we eliminate small regions and perform dilation processes to the size of the tumor little bit larger, and **3) Estimation of the major axis** where we defined the major axis as the segment that satisfies the condition of passing through two points on the contour line of the tumor area and the centroid of that. Using this process, the tumor is rotated so as to align the determined major axis along the horizontal direction.

B. CNN classifier for evaluation of the proposed pre-processing method

In order to evaluate the effect of the proposed pre-processing method, we designed a simple CNN-based melanoma classifier, implemented using the Caffe framework [15]. We emphasize here that the objective of this paper is the proposition of a simple and effective pre-processing technique for melanoma discrimination and its objective evaluation rather than maximizing the classification performance on a specific dataset. The benefit of using a CNN in this study is, of course its excellent learning capability, but more importantly, it eliminates the manual parameter design that may bias results. In fact, introducing manually designed parameters associated with the symmetry of the tumor for conventional classifiers such as support vector machines can disrupt object evaluation.

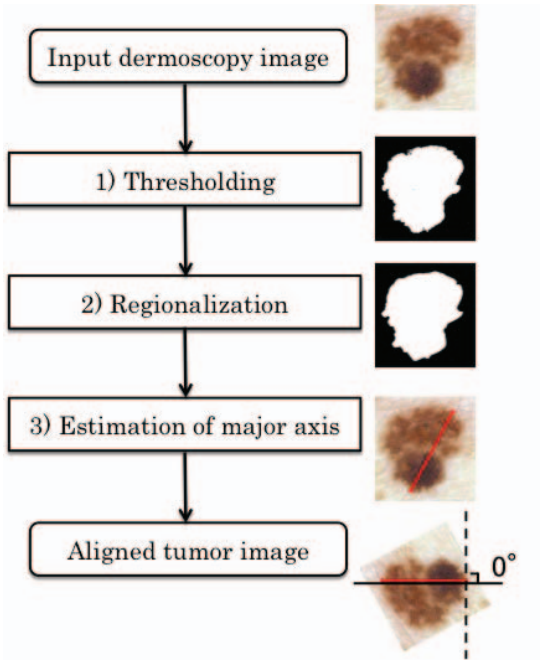

Fig. 1. Flowchart of the major axis alignment process.

Our CNN-based classifier is a simplified version of the so-called AlexNet [9] and is composed of an input layer, three convolution parts, each of which has convolution, pooling, and local contrast normalization (LCN) layers, and an output layer, as shown in Fig. 2. The structure and parameters of our CNN are determined based on our preliminary experiments.

The first convolutional part (Layer 1) receives 224 × 224 × 3 patches, which are randomly extracted from 256 × 256 input images with 96 kernels. The size of these kernels is 11 × 11 × 3 and stride size is two pixels. The second convolutional part (Layer 2) takes the output of Layer 1 as the input and filters that output with 48 kernels (size: 7 × 7 × 96, stride = 2). The third convolutional part (Layer 3) has 128 kernels (size: 5 × 5 × 48, stride = 2) connected to the output of Layer 2. Layers 2 and 3 take the subsampled and normalized output, respectively, of the previous layer as their input. All the pooling parts summarize a 3 × 3 neighborhood (i.e., max-pooling) with a stride of two pixels. Finally, the output layer is connected to all the outputs of the previous layer and fed to a two-way softmax function, which produces a probability distribution over

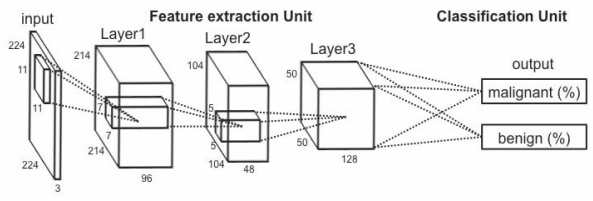

Fig. 2. Architecture of our simple CNN.

the two-class labels (i.e., malignant or benign).

CNNs automatically generate effective internal patterns for classification that respond to neighboring patterns such as large gradients and line segments with specific directions by means of their convolutional training process. Because nevi and melanomas develop differently from a medical point of view, we expect that the alignment of the major axis of the tumor will help in identifying features that assist their classification.

III. Experiments and Results

The image set used in this study consists of 1,760 dermoscopic images (including 329 cases of melanomas and 1,431 cases of nevi) acquired from three institutions (University Federico II of Naples, Italy; University of Graz, Austria; University of Vienna, Austria) and an open dataset (International Skin Imaging Collaboration). Regarding the data from the three institutions, lesions were diagnosed based on histopathological examination of biopsies. In each dermoscopic image, the boundary box of the tumor was roughly determined based on the preceding method. Each dermoscopy image was then cropped in a square shape. The length of the sides of each image was equal to that of the long side of the boundary box. The cropped images were resized to 256 $\times$ 256 pixels using bilinear interpolation. Note that in each training epoch and evaluation, each sub-image having 224 $\times$ 224 pixels (i.e., matching the size of the first convolutional part) was randomly cropped from the given image (256 $\times$ 256 pixels) and input to the CNN.

We conducted three evaluation experiments to confirm the effectiveness of the proposed alignment of the major axis of the tumor:

- Experiment A (performance baseline): The image set was artificially augmented four times by generating horizontal and vertical inversions (1,760 $\times$ 4 = 7,040 images).
- Experiment B (data augmentation): The image set was artificially augmented 36 times by rotating the image in increments of 10 degrees (1,760 $\times$ 36 = 63,360 images).
- Experiment C (proposed): The images were aligned with the major axis of the tumors in a horizontal direction and then artificially augmented four times by generating horizontal and vertical inversions (1,760 $\times$ 4 = 7,040 images).

Note that the only difference between Experiments A and C was the use of the proposed image alignment method; the number of images trained by each CNN was the same. The augmented dataset in Experiment B included similar images obtained by the proposed image alignment.

We used a cross entropy objective function as the optimization target and the stochastic gradient descent method with mini-batches of 100 images to optimize the trainable parameters of the CNN. The CNN was trained 100 epochs, and weights were updated at the end of each mini-batch process. Sensitivity and specificity were used as evaluation criteria. The performance of the classification accuracy of our CNN was evaluated using a five-fold cross validation test.

TABLE I
Melanoma classification results for the three experiments

	SE (%)	SP (%)	AUC	# training images
Experiment A	74.9	80.9	0.789	1,760×4×(4/5)
Experiment B	77.9	88.8	0.826	1,760×36×(4/5)
Experiment C	80.9	88.1	0.847	1,760×4×(4/5)

A summary of the obtained results from the experiments is shown in Table I, in terms of sensitivity, specificity and the area under the receiver-operating-characteristic (ROC) curve. Fig. 3 shows the ROC curves derived from the classification results.

IV. Discussion

As can be seen, the CNN with the proposed tumor alignment procedure (Experiment C) attained superior classification performance (84.7% in terms of AUC) compared both Experiments A and B. While a classification performance of 84.7% AUC might not be extremely high, as mentioned above, the objective of this research is to propose a simple yet effective pre-processing method for melanoma classification based on cytological aspects, and therefore we need to evaluate its effect with the minimum configuration and eliminate other factors as much as possible.

We want to emphasize that the benefit of the proposed tumor alignment was greater by nine-fold that of data augmentation. This means that the asymmetrical characteristics of melanoma were effectively used in the CNN and helped to improve classification accuracy. Here, the augmented data used in Experiment B included images with the major axis aligned in a nearly horizontal direction. The fact that only the method with effective alignment showed superior performance is interesting to note. The required time for training CNN depended considerably on the size and number of training images and training epochs. Extension of training data to include image rotation actually improved classification accuracy, even though longer

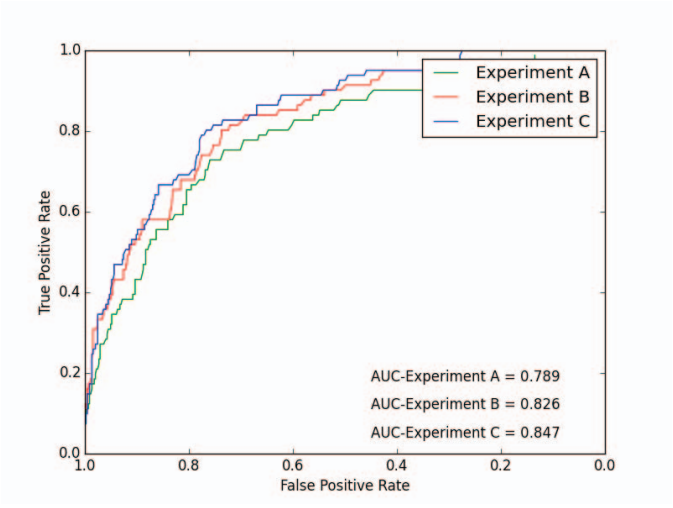

Fig. 3. ROC curves for different pre-processing methods

training time was then required. This issue cannot be ignored when the training data is large. The simple and effective tumor alignment strategy helped to improve both accuracy and computational cost.

In this study, we estimated the major axis of the tumor based on a very simplified tumor extraction method, because more accurate estimation often requires complicated procedures and high computational cost as shown in previous studies. Employing a more sophisticated method to estimate the major axis might thus lead to improved performance. However, as previously mentioned, a trade-off exists. The proposed pre-processing can be applied easily for state-of-the-art CNN-based classification methods (i.e. use very large scale pre-trained CNN models as a feature extractor as in [11]), and we plan to report in this in the near future.

V. Conclusions

In this paper, we have proposed a simple and effective pre-processing method for automated melanoma classification, in which we align the major axis of the tumor prior to training. Based on an extensive set of experiments, we have confirmed that our proposed pre-processing technique improves considerably the classification performance of melanoma.

Appendix

ABCD rule [13] : The ABCD rule is one of the most well-known and commonly used semi-quantitative diagnosis schemes for melanoma. It quantifies the asymmetry (A), border sharpness (B), color variegation (C) and the number of differential structures (D) present in a lesion. Using the ABCD rule, physicians determine a score in each category ($0 \leq A \leq 2$, $0 \leq B \leq 8$, $1 \leq C \leq 6$, and $1 \leq D \leq 6$) and the total dermoscopy score (TDS) of a lesion is calculated as TDS = $(A \times 1.3) + (B \times 0.1) + (C \times 0.5) + (D \times 0.5)$. A TDS below 4.75 indicates benignity, whereas one above 5.45 indicates malignancy. Our pre-processing method primarily focuses on the symmetry of the lesion corresponding to the 'A' component of the ABCD rule. According to the diagnosis by 40 expert dermatologists using the ABCD rule, the diagnostic accuracy was 83% in terms of sensitivity and less than 73% in terms of specificity [5].

Acknowledgments

This research was partially supported by the Ministry of Education, Culture, Science and Technology of Japan (Grant in-Aid for Fundamental Research Program (C), 26461666, 2014-2017).

References

[1] Australian Institute of Health and Welfare, "Australian Cancer Incidence and Mortality books 2015" ?http://www.aihw.gov.au/acim-books/?, Jan. 2015.

[2] H. P. Soyer, J. Smolle, H. Kerl, and H. Stettner, "Early diagnosis of malignant melanoma by surface microscopy," Lancet, Vol. 2, No. 8562, p. 803, Oct. 1987.

[3] J. Mayer, "Systematic review of the diagnostic accuracy of dermoscopy in detecting malignant melanoma," Med. Journal of Australia, Vol. 167, No. 4, pp. 206-210, Aug. 1997.

[4] W. Stolz, O. B. Falco, P. Bilek, M. Landthaler, W. H. C. Burgdorf, and A. B. Cognetta, "Color Atlas of Dermatoscopy – 2nd enlarged and completely revised edition," Berlin, Blackwell Publishing, 2002.

[5] G. Argenziano, H. P. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sara et al., "Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet," Journal of American Academy of Dermatology, Vol. 48, No. 5, pp. 679-693, May. 2003.

[6] M. E. Celebi, H. Iyatomi, W. V. Stoecker, R. H. Moss, H. S. Rabinovitz, G. Argenziano, and H. P. Soyer, "Automatic Detection of Blue-White Veil and Related Structures in Dermoscopy Images," Computerized Medical Imaging and Graphics, Vol. 32, No. 8, pp. 670-677, 2008.

[7] Z. Liu, J. Sun, L. Smith, M. Smith, and R. Warr, "Distribution quantification on dermoscopy images for computer-assisted diagnosis of cutaneous melanomas," Medical & biological engineering & computing, Vol. 50, No. 5, pp. 503-513, 2012.

[8] K. Shimizu, H. Iyatomi, M. E. Celebi, K. Norton, and M. Tanaka, "Four-class classification of skin lesions with task decomposition strategy," IEEE Trans. Biomedical Engineering, Vol. 62, No. 1, pp. 274-283, 2015.

[9] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems, pp. 1106-1114, 2012.

[10] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," in MICCAI MLMI, Vol. 9352, pp. 118126, 2015.

[11] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep Features to Classify Skin Lesions," IEEE 13th International Symposium on Biomedical Imaging, pp. 1397-1400, 2016.

[12] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks," ICLR, arXiv:1312.6229v4, 2014.

[13] W. Stolz, A. Riemann, A. B. Cognetta, L. Pillet, W. Abmayr, and D. Holzel, "ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma," European Journal of Dermatology, No. 4, No. 7, pp. 521-527, Apr. 1994.

[14] H. Iyatomi, H. Oka, M. E. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka et al., "An improved Internet- based melanoma screening system with dermatologist-like tumor area extraction algorithm," Computerized Medical Imaging and Graphics, Vol. 32, No. 7, pp. 566-579, 2008.

[15] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick et al., "Caffe: Convolutional Architecture for Fast Feature Embedding," arXiv preprint arXiv: 1408.5093, 2014.