Intelligent Systems Conference 2018
6-7 September 2018 | London, UK

The impact of replacing complex hand-crafted features with standard features for melanoma classification using both hand-crafted and deep features

Binu Melit Devassy (*First Author*)
Computer science
NTNU Norwegian University of Science and Technology
Gjovik, Norway
binum@stud.ntnu.no

Jon Yngve Hardeberg (*Third Author*)
Computer science
NTNU Norwegian University of Science and Technology
Gjovik, Norway
jon.hardeberg@ntnu.no

Sule Yildirim-Yayilgan (*Second Author*)
Computer science
NTNU Norwegian University of Science and Technology
Gjovik, Norway
sule.yildirim@ntnu.no

Authors Name/s per 4th Affiliation (*Fourth Author*)
line 1 (of *Affiliation*): dept. name of organization
line 2: name of organization, acronyms acceptable
line 3: City, Country
line 4: e-mail address if desired

***Abstract*— Melanoma is the deadliest form of skin cancer and it is the most rapidly spreading cancer in the world. An earlier detection of this kind of cancer is curable, hence earlier detection of melanoma is pre-eminent. Because of this fact, a lot of research is being done in this area especially in automatic detection of melanoma. In this paper, we are proposing an automatic melanoma detection system which utilize a combination of deep and hand-crafted features. We analyzed the impact of using a simpler and standard hand-crafted feature, in place of complex usual hand-crafted features eg. shape, texture, diameter, or some custom features. We used a convolutional neural network (CNN) known as deep residual network (ResNet) to extract the deep features and utilized the scale invariant feature descriptor (SIFT) as the hand-crafted feature. The experiments revealed that, combining SIFT did not improve the accuracy of the system, however we obtained higher accuracy than state of the art methods with our deep only solution.**

***Keywords*— melanoma detection; ResNet; SIFT;**

I. Introduction

Melanoma, the treacherous skin cancer is primarily caused by UV radiation from sun or other sources. Around 10,130 people died annually in US because of Melanoma [1]. In Norway, incidence rate (number of cases per 100,000 per year) is dramatically increased from 1.9 and 2.2, for women and men respectively, to 19.6 and 19.0, over a period of 57 years from 1953 to 2010 [2]. The major cause for skin cancer is determined as the exposure to intense radiations from sun. Therefore, the primary prevention should be avoiding long term exposure to solar radiation or use some protection against sun exposure. Skin cancer could have been broadly classified as melanoma (Fig 1a) and non-melanoma (Fig 1b). Non-melanoma cancers are more common and have a higher chance of recovery than melanoma, because they are less capable of spreading from one part to the other parts of the body [3]. An early detection of melanoma cancer is almost curable, if not it can be fatal. The most common diagnosis method is visual inspection with dermoscopy followed by histopathological examination as required [4]. The ABCD(E) rule is the widely accepted technique for clinical inspection of skin lesions [5].

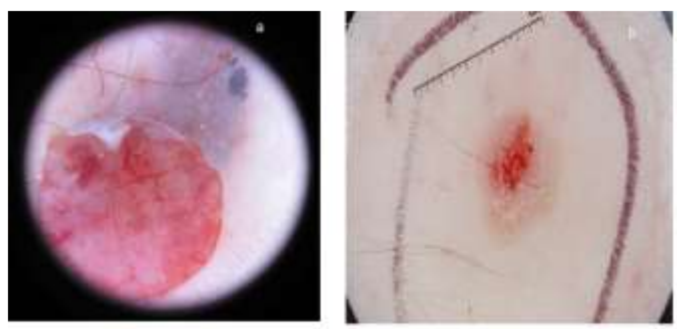

Figure 1 Examples of skin lesions (a)) melanoma and (b) seborrheic keratosis (images taken from ISIC challenge 2017) [6]

The detection of melanoma is still continued as a difficult problem. The manual inspection is considered as the effective solution, but it is also difficult and time consuming, because most of the people have many non-cancerous lesions on their skin. Due this fact and its importance, many researchers are working on automatic and semi-automatic detection of melanoma. The recent trend in this field is a hybrid classifier that employs both deep and hand-crafted features for detecting melanoma. Most of these hybrid classifiers are using computationally expensive complex features as hand-crafted features [7].

In this paper we are comparing the degradation or improvement in accuracy while replacing the complex features with simple well-known features in hybrid melanoma classifiers. We compared our result with the result from the ISIC 2017 Melanoma Detection challenge [6]. Here we are interested in detecting melanoma and not interested in other classes of non-melanoma skin lesions.

This paper is organized as follows. In section 2, will discuss about the related works. In section 3, proposed method will be explained. In Section 4, we will discuss the details about the experiments, in section 5, we will analyses the results obtained. The section 6 will discuss about the contributions into this project and section 7 has the conclusions and future work

II. RELATED WORK

We can find many examples of work related to combining deep features with hand crafted features, consider the work done by Raul Paul et al [8]. They obtained an accuracy of 90% in predicting lung cancer by combining deep and hand-crafted features, initially they had 77.5% accuracy from hand crafted features and 77.5% from deep features. Here [9] is another example of hybrid implementation for chest pathology detection. Bram van Ginneken et al [10] also used a combination of CNN and hand-crafted features to detect Pulmonary Nodule detection and obtained a significantly better result.

This work is different from the predecessor because of the simplicity of the hand-crafted feature that we have used. As mentioned earlier we have used a combination of hand-crafted and deep features. We used SIFT (scale invariant feature transform) [11] a well-known feature descriptor as hand crafted feature and used ResNet [12] for extracting deep features.

We have used the data set provided by the ISIC challenge 2017 for training and evaluation to make a reasonable comparison with state of the art results. The data set contains total 2000 skin lesion images, contains 374 melanoma images and 254 seborrheic keratosis images. We have done data augmentation to increase the robustness of the training, and we will discuss about the process flow of our approach and the experiments in the upcoming sections.

III. THE PROPOSED METHOD

Image pre-processing, segmentation, feature extraction and classification are the basic building blocks of an algorithm to detect a skin lesions [3]. In our proposed method (Fig 2), we introduced data augmentation to increase the robustness of the training. The input image will be passed through the hand-crafted and deep feature extraction methods and both features will be combined in the form of a single vector and passed into the classifier. We will describe each block in the detail in following sessions.

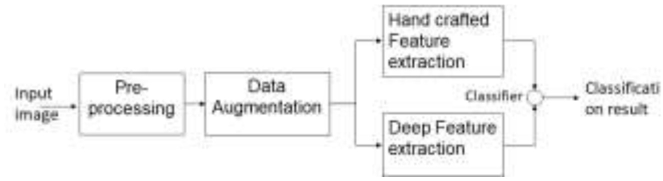

Figure 2 High level diagram of the proposed framework

A. Pre-processing

Currently there is no standard protocol available for capturing and transmitting skin lesion images. Skin lesion images are coming from entirely different set of conditions like varying illuminance, different capturing devices and different angles of capturing. Here we are more concerned about the changes in the illuminant, it will affect the color appearance of the lesion images and thus the accuracy of the system. To overcome this, we are applying color constancy algorithms on the input images. There are several approaches that are available in the literature like Grey world, Max_RGB, Shades of Gray etc. [13], and we obtained promising results with the Gray world method.

B. Data augmentation

It is a good practice to do data augmentation to improve the performance of deep neural networks especially when we have a small amount of training data [14]. The data augmentation can improve the robustness of the deep neural network. The common methods for data augmentation include various geometric transformation, cropping and flipping. In our approach we are more interested in orientation changes and cropping, because as we mentioned earlier there is no protocol available for capturing and transmitting the skin lesion images, so different clinicians will use different angles for capturing the image. We have used a combination of two crops and four rotations (45, 90, 135 and 180 degrees), hence for each input image we have generated eight augmented versions. The cropping rectangles are the largest inner rectangles ensuring that all pixels belong to the original image. Finally, we performed a scaling to the square size of 225 x 225 because our CNN requires a square input image.

C. Feature Extraction

Feature extraction is the core part of any classification problem. As we mentioned earlier most of the hybrid-methods are using a complex hand-crafted feature with deep network. In our approach we are using SIFT descriptor as our hand-crafted feature and ResNet for extracting deep features. We have selected SIFT because, it is a proven and well known descriptor in computer vision applications, which is invariant to most of the image transformations [15]. Another fact behind choosing SIFT as a candidate is the computation cost, SIFT algorithm is already used in many real-time applications [16][17]. However, we do not have any computational time matrix related to the other hand-crafted features for a comparison. The most important fact is the hybrid approach with SIFT and deep is unique in melanoma detection.

While using SIFT as a feature set, we faced the problem of feature set reduction, it was a challenging problem because each SIFT descriptor has 128 features and for each image we

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

got an average 125 SIFT descriptors. We approached this problem with two different methods for feature set reduction, the first approach was using a SIFT match to obtain the best features and in the second approach we have used the bag of words method for feature set reduction. In the first method, we used a SIFT matching [18] of the original image with its corresponding augmented images to obtain the matching descriptor. From the matching descriptors, we have chosen top twenty features based on their scores. In the second approach, we have used the standard bag of word using k-means [19]. We grouped the descriptors into different number bags (10, 20, 50 and 100) and we found that twenty bags are giving the best performance while combining with deep features.

In deep learning, more network depth is a desirable feature, but deeper networks are difficult to train. The major problem with a deeper network is the accuracy saturation and then a rapid degradation, and adding more layers will end up with a higher training error [12].

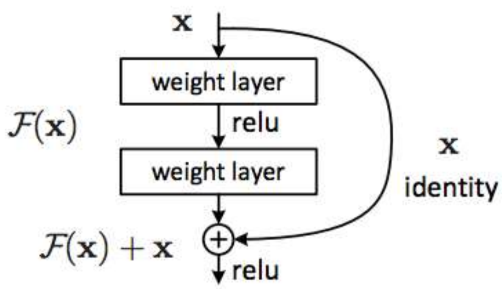

Figure 3 Fundamental building block of a residual network[20]

The deep residual network also known as ResNet is made up of building blocks known as residual blocks as shown in Fig 3 and each block contains several convolutional layers, batch normalization layer and ReLU layers. The residual block will allow to bypass few convolution layers at a time [20]. Therefore, ResNet is capable of overcoming the limitation of other deep networks by adding the shortcut connections. Hence, we decided to choose the ResNet for extracting deep features. We have used a pretrained ResNet model to perform the classification, which was generated using the ImageNet ILSVRC challenge data [20]. The last layer of this network has 2048 features extracted from the input image and we tap these features for doing the classification.

D. Classification

Now we have the features set from both the deep network and the SIFT descriptors, here we will combine both of the features and feed into the classifier. Our case is a binary classification, melanoma or not. We explored SVM and RusBoost [21] algorithms to generate the classifier. We will discuss the performance difference between these two in the results section.

E. Comparison with State of the art

In this section we are doing a comparison of state of the art method [22] against ours. In general, we can say our approach is a canonical approach with all necessary modules while the other method has some extra modules for additional processing. The major differences are:

- We are using a standalone CNN while the other method is based on an ensemble of CNNs.
- The authors are using age and sex information's with the features extracted from the input image for prediction, but we are using only the input image.
- They are using more complex pre-processing, we are using relatively simple pre-processing.

IV. Experiments

We have used MatConvNet tool box [23] for extracting the deep features and used vlfeet SIFT library [24] for extracting the SIFT descriptors. We performed a number of experiments to understand the impact of combining SIFT features with deep features. For that, first, we performed training and evaluation only using the deep features, then only using the SIFT features and finally we have done the hybrid evaluation using both the deep and the SIFT descriptors. We have also done all the above experiments with both SVM and RusBoost classifiers. Initially we had 2000 images that are augmented into 16000 images using the combinations of two crops and 4 orientations. We divided the entire data set into ten groups and performed a cross validation by using nine groups for the training and one for the evaluation.

V. Results and Evaluation

The classification results are validated based on the accuracy values. The table 1 summarizes the results for deep features plus SIFT features, deep features only, SIFT features only and state of the art result. We can observe that the result from deep only features with the SVM classifier is giving the highest accuracy and SIFT only features with the RusBoost classifier is giving the lowest among all. From the result it is clear that, combining SIFT with deep features are decreasing the accuracy around 2% in the SVM classifier and remains unchanged with the RusBoost classifier. We can derive the reason for this behavior from the SIFT only and deep only results where the accuracies are 0.7849 and 0.8324 respectively as follows: while combining both, we are getting an accuracy approximately equal to the average of both values 0.8041. In the case of RusBoost classifier, we are getting the result (0.821) almost same as the highest (0.8266) in the combination, which is the deep feature only, instead of the average value. The important fact is that with the deep only features we achieved an accuracy (0.8324) greater than that of the state of the art method (0.828).

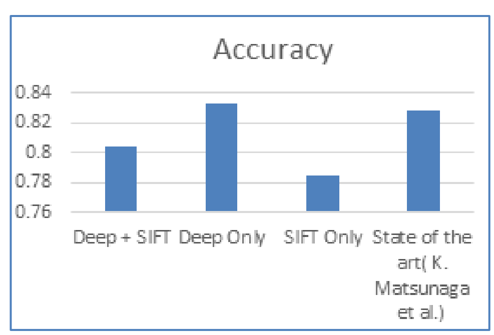

Figure 4 The result comparison , we can see that our deep only method has more accuracy than the state of the art

We also measured the speed of our processing for a single image, and the deep only approach is taking 0.372 seconds average processing time for evaluating and with the SIFT it is taking 0.7945 seconds in the average. These are the measurements taken using Matlab R2017_a, on a system with a 12 core Intel Xeon processor with 64GB RAM without using GPU acceleration.

VI. CONTRIBUTION

Through this experiment we want to analyze that, whether combining the SIFT features with the deep features will make any improvement in the final result. We are also interested in comparing our results with state of the art results. We achieved this using the limited number of data from the ISIC challenge 2017 using data augmentation, ResNet, and SIFT with SVM and RUSBoost classifiers.

	Deep + SIFT	Deep Only	SIFT Only	State of the art(K. Matsunaga et al.)
SVM	Sn: 0.3435	Sn: 0.3063	Sn: 0.1029	NA
	Sp: 0.9113	Sp: 0.9568	Sp: 0.9434	Sp: 0.851
	Acc: 0.8041	Acc: 0.8324	Acc: 0.7849	Acc: 0.828
	Pr: 0.4592	Pr: 0.6022	Pr: 0.2989	NA
	Rc: 0.3435	Rc: 0.3063	Rc: 0.1029	NA
	MCC: 0.2811	MCC: 0.3442	MCC: 0.0738	NA
	Fs: 0.3812	Fs: 0.3972	Fs: 0.1450	NA
RUSBoost	Sn: 0.4536	Sn: 0.4408	Sn: 0.6364	NA
	Sp: 0.9080	Sp: 0.9182	Sp: 0.5981	NA
	Acc: 0.8217	Acc: 0.8266	Acc: 0.6044	NA
	Pr: 0.5137	Pr: 0.5321	Pr: 0.2383	NA
	Rc: 0.4536	Rc: 0.4408	Rc: 0.6364	NA
	MCC: 0.3723	MCC: 0.3784	MCC: 0.1753	NA
	Fs: 0.4681	Fs: 0.4678	Fs: 0.3467	NA

Table 1 Average results after cross-validation, using 20 BOW SIFT

VII. CONCLUSION AND FUTURE WORK

From the results discussed above, we can come to the main conclusion that combining SIFT with deep feature did not improve the results. For the case of the SVM, the accuracy reduced around 2% after combining the deep features with the SIFT features. We are considering this work as an initial mile stone in this area where a combination of the SIFT features and deep features are applicable. However still we can see some unexplored area where future researchers can work: the first area is the feature reduction method. We tried only two, still there exists few other methods to explore. The other opportunity is to explore the various feature selection methods and their combinations [25]. Also, we can explore other classifiers like Decision Trees [26], Random Forest [27], Symmetric Uncertainty Feature Selector [28] etc.

We can observe the processing time is very small and we can improve it by using low level programming languages(C++) with GPU acceleration. We could not make a comparison of processing time because of unavailability of a standard processing time. We believe that this could be another opportunity as future work, to improve the speed along with accuracy to achieve a real-time scanning.

The next important finding is that, the performance of our deep only feature with SVM classifier out performs state of the art. We achieved this by using rather simpler canonical approach than a complex approach. They also used age and sex information along with the images, that is another future work related to this proposal. Moreover, we need a larger data set to do a perfect training and evaluation, because all the available deep networks are trained with approximately one million images from ImageNet [29].

We are not doing any segmentation of the skin lesion as a part of the pre-processing step. We believe that our results will improve if we can integrate some reliable segmentation method as stated in [30]. Some studies show that pattern recognition yields better results over other approaches [31], hence the researches in future can emphasize on some other standard and simpler hand-crafted features to uncover the pattern from the skin lesion images instead of the SIFT descriptor.

REFERENCES

[1] "Melanoma." [Online]. Available: http://www.skincancer.org/skin-cancer-information/melanoma. [Accessed: 16-Oct-2017].

[2] I. K. Larsen, "Cancer in Norway," 2010.

[3] R. B. Oliveira, M. E. Filho, Z. Ma, J. P. Papa, A. S. Pereira, and J. M. R. S. Tavares, "Computational methods for the image segmentation of pigmented skin lesions: A review," *Comput. Methods Programs Biomed.*, vol. 131, pp. 127–141, 2016.

[4] E. M. Wurm and H. Peter Soyer, "Scanning for melanoma," *Aust. Prescr.*, vol. 33, no. 5, pp. 150–155, 2010.

[5] H. M. Shaw, D. S. Rigel, R. J. Friedman, W. H. Mccarthy, and A. W. Kopf, "Early Diagnosis of Cutaneous Melanoma Revisiting the ABCD Criteria," vol. 292, no. 22, pp. 2771–2776, 2004.

[6] "ISIC 2017: Skin Lesion Analysis Towards Melanoma Detection." [Online]. Available: https://challenge.kitware.com/#challenge/n/ISIC_2017%3A_Skin_Lesion_Analysis_Towards_Melanoma_Detection. [Accessed: 28-Nov-2017].

[7] T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "Combining

deep learning and hand-crafted features for skin lesion classification," *2016 6th Int. Conf. Image Process. Theory, Tools Appl. IPTA 2016*, 2017.

[8] R. Paul *et al.*, "Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma," *Tomogr. a J. imaging Res.*, vol. 2, no. 4, pp. 388–395, Dec. 2016.

[9] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 294–297.

[10] B. van Ginneken, A. A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 286–289.

[11] D. G. Lowe, "Object recognition from local scale-invariant features," *Proc. Seventh IEEE Int. Conf. Comput. Vis.*, pp. 1150–1157 vol.2, 1999.

[12] S. Wu, S. Zhong, and Y. Liu, "Deep residual learning for image steganalysis," *Multimed. Tools Appl.*, pp. 1–17, 2017.

[13] M. E. Celebi, T. Mendonca, and J. S. Marques, "Dermoscopy Image Analysis [Book review]," *IEEE Trans. Med. Imaging*, vol. 35, no. 4, pp. 1147–1148, Apr. 2016.

[14] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 1, pp. 31–40, Jan. 2017.

[15] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 2004, vol. 2, p. II-506-II-513 Vol.2.

[16] N. H. Dardas and N. D. Georganas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 11, pp. 3592–3607, 2011.

[17] M. Lalonde, D. Byrns, L. Gagnon, N. Teasdale, and D. Laurendeau, "Real-time eye blink detection with GPU-based SIFT tracking," in *Computer and Robot Vision, 2007. CRV '07. Fourth Canadian Conference on*, 2007, pp. 481–487.

[18] D. G. Lowe, "Distinctive Image Features from," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[19] C. S. Venegas-Barrera and J. Manjarrez, "Visual Categorization with Bags of Keypoints," *Rev. Mex. Biodivers.*, vol. 82, no. 1, pp. 179–191, 2011.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[21] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A Hybrid Approach to Alleviating Class Imbalance," *IEEE Trans. Syst. Man, Cybern. - Part A Syst. Humans*, vol. 40, no. 1, pp. 185–197, Jan. 2010.

[22] K. Matsunaga, A. Hamada, A. Minagawa, and H. Koga, "Image Classification of Melanoma, Nevus and Seborrheic Keratosis by Deep Neural Network Ensemble," *1 Isic 2017*, pp. 2–5, 2017.

[23] A. Vedaldi and K. Lenc, "MatConvNet - Convolutional Neural Networks for MATLAB," in *Proceeding of the {ACM} Int. Conf. on Multimedia*, 2015.

[24] "SIFT detector and descriptor." [Online]. Available: http://www.vlfeat.org/overview/sift.html. [Accessed: 29-Nov-2017].

[25] Y. Chen and C. Lin, "Combining SVMs with Various Feature Selection Strategies," *Featur. Extr.*, vol. 324, no. 1, pp. 315–324, 2006.

[26] J. R. Quinlan, "Induction of Decision Trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

[27] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, 1998.

[28] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution," *Int. Conf. Mach. Learn.*, pp. 1–8, 2003.

[29] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[30] N. Codella *et al.*, "Deep Learning Ensembles for Melanoma Recognition in Dermoscopy Images," vol. 61, no. 4, pp. 1–28, 2016.

[31] 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.