HIGH SCHOOL EDITION
Journal of Student Research
Volume 11 Issue 3 (2022)

Melanoma Skin Cancer Classification via Region-Aware Hierarchical Feature Aggregation

Kyungryun Kim[1]

[1]Detroit Country Day School

ABSTRACT

Melanoma is a type of skin cancer with the highest risk of death. It is critical to identify these early, as the chances of survival significantly drop after stage 2 melanoma. The use of many technologies has lessened this risk but is still limited when correctly identifying malignant ones. Classifying malignant skin cancers is difficult for the following reasons: the shapes and sizes of melanomas are irregular. It is also challenging to visually distinguish between melanomas and non-melanoma regions. The performance of the previous research is based on the depth of the networks. This has a significant trade-off as using more layers adds to the computational cost of the process. Also, their methods use melanoma segmentation as an auxiliary input to the classification network. Thus, the performance of the classification significantly drops when the segmentation task fails. To address this issue, I propose a novel melanoma classification network that uses hierarchical feature aggregation with an attention mechanism. The overall architecture of the proposed network is as follows: The melanoma feature extractor takes the melanoma image as input and produces the image feature related to melanoma. The second module, the attention network, takes the same melanoma image and outputs the attention map which provides the melanoma feature extractor with feature-level regions of interest. The proposed network achieves an accuracy of 82.7% on the melanoma detection dataset which is publicly available online. Throughout the experiments, I have shown that the proposed method outperforms the previous state-of-the-art methods.

Introduction

Melanoma is a type of skin cancer that causes the layers of skin tissues to develop moles. This is caused by pigment-producing cells called melanocytes. It is often difficult to differentiate between the types of Melanomas. The techniques commonly used today to treat and detect these diseases are dermoscopy and different types of biopsies. Melanoma skin lesions are often examined by using dermoscopy. Dermoscopy uses an instrument called a dermatoscope to examine skin tissues at the microscopic level. It is critical to detect skin cancer in the early stages. Recently, a study has found that about 44.27 percent of dermatologists use the Dermoscopy technique to detect skin lesions and the other half use biopsies. These techniques have been proven to be useful in detecting skin cancer. However, these techniques carry a downside. This is because the lack of resolution limits dermoscopy's accuracy in detecting melanoma types. For the biopsy, its downside is that there might be an infection or different complications.

Much research has been conducted to improve several weaknesses of certain methods of biopsies and in dermoscopy. Codella et al. (Codella, Noel, et al 2019) proposed that machine learning-based multiclass probability outperformed content-based image retrieval. Subsequently, Gomez published another study et al. (Argenziano, Giuseppe, et al. 2012) and they used an independent histogram pursuit, which is an unsupervised algorithm for skin lesion segmentation. Additionally, another researcher, Adekanmi, applied deep learning for skin lesion detection by developing a type of fully convolutional residual network (Adegun, Adekanmi A., and Serestina Viriri. 2019). Using a new method called the lesion classifier, Adekanmi and Serestina were able to establish an accuracy of 95%.

As aforementioned, numerous research has shown that applying the deep learning method for detecting skin cancer is feasible and that it shows comparable performance. However, there are still many challenges that remain.

One of those challenges is that there can be a different contrast between the normal skin and the cancerous area. This contrast often degrades the performance of the trained classifier model. In addition, it is difficult to distinguish between malignant types and benign ones since some are visually like each other.

To address this problem, I propose a novel melanoma classifier system. In this paper, I consider melanoma categorization as a fine-grained classification since there are a lot of various categories of melanoma that are visually similar. The proposed classifier system is composed of a hierarchical melanoma feature extractor, an Attention module, and a linear Layer. The proposed hierarchical feature extractor takes three levels of pyramid images as inputs and outputs feature maps corresponding to each pyramid level. The Attention module takes the top-level of the pyramid image as an input and generates an attention map as an output. The attention map represents the region of interest that is expected to be the melanoma area. Then, the attention map and the concatenated multi-scale feature map are multiplied pixel-wisely. This is then fed to the linear layer where it produces the final prediction. I also proposed a novel skin color jitter augmentation to make the model robustly perform against input variants. The proposed method achieves accuracy of 82.7% on publicly available melanoma detection dataset.

Methods

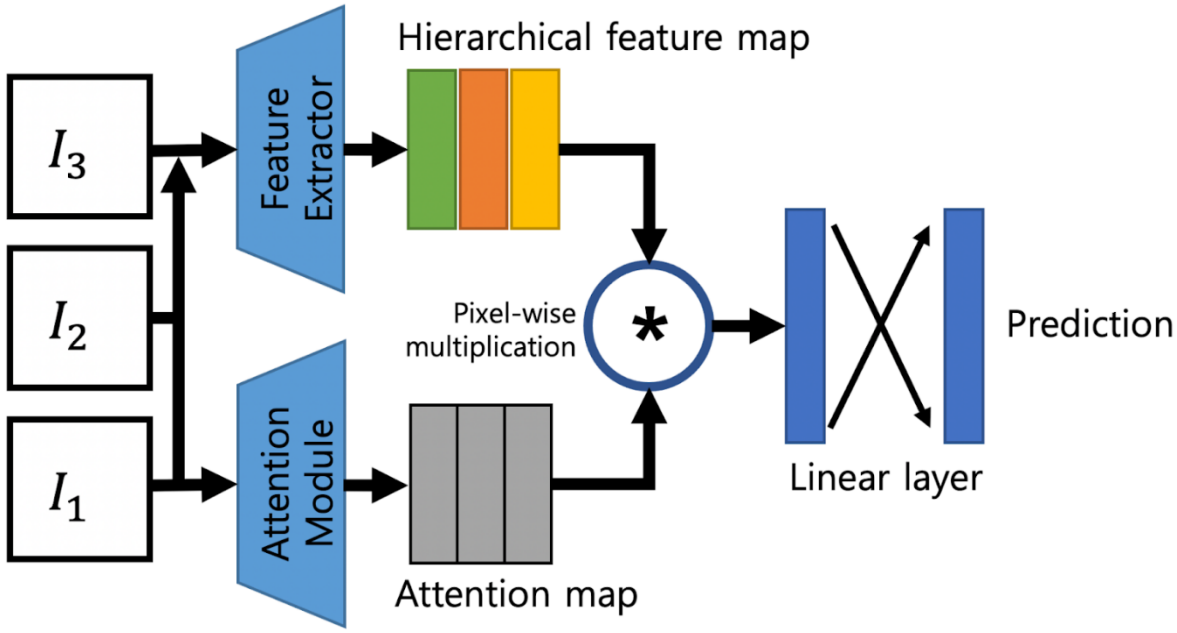


Figure 1. The overall architecture of the proposed melanoma classification system.

Fig. 1 represents the architecture of the proposed melanoma classification system. The proposed system is composed of three modules. The first module hierarchical melanoma feature extractor takes three levels of pyramid melanoma images as input and outputs the hierarchical image features. Given the same input image, the attention module generates an attention map that focuses on the melanoma region by displaying the melanoma-related pixel area in the input image. The generated attention map is then applied to the image features by performing pixel-wise multiplication. Finally, the feature map is fed to the Linear layer where the score vector of melanoma categories is calculated. In chapter 2.1, I explain how the proposed hierarchical melanoma feature extract operates. I also explain how I apply the attention module to the proposed system in chapter 2.2 in detail. Finally, I explain how I implement the entire system in chapter 2.3.

Hierarchical Melanoma Feature Extractor

Given an input melanoma image I_1, the proposed system aims to predict the type of melanoma $P = \{P_1, P_2, P_3\}$. Where P_1, P_2, and P_3 denote Seborrheic keratosis, Melanoma, and Nevus, respectively. The proposed melanoma feature extractor ***MFE*** takes three different levels of pyramid image $I = \{I_1, I_2, I_3\}$. Fig. 2 shows an example of how the pyramid image generation is produced by receiving an input image of melanoma and down sampling it two times by ½.

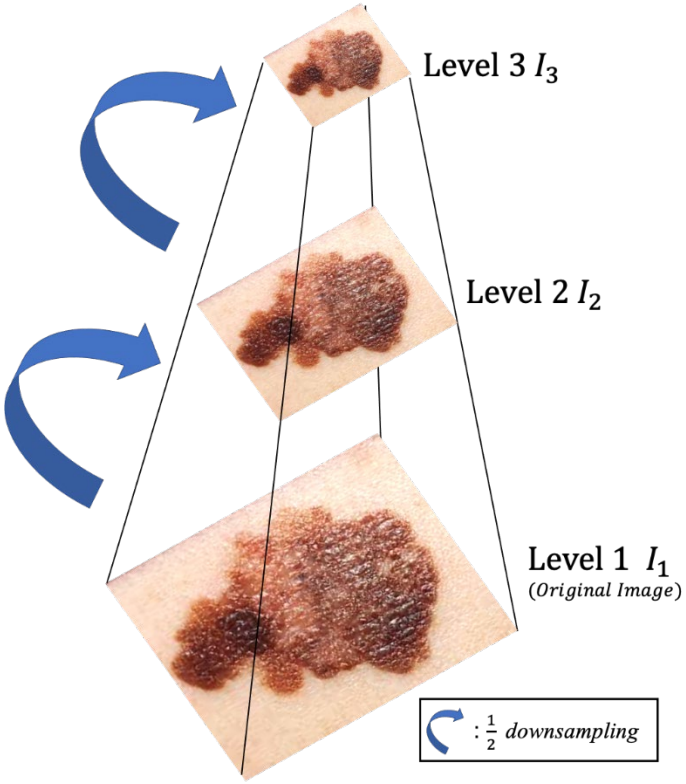


Figure 2. Example of the pyramid image used in this paper.

With given pyramid image I, the proposed ***MFE*** produce a hierarchical feature map $F = \{f_1, f_2, f_3\}$. Where I_2 denotes the downsizing of I_1 by a factor of $\frac{1}{2}$ and I_3 denotes the downsizing of I_1 by a factor of $\frac{1}{4}$. I first aggregate this hierarchical feature maps into one feature map. This feature map contains rich information on melanoma images as it is extracted from different resolution images. The hierarchical feature map F proceeds through a pixel-wise multiplication with the attention map and is then fed to the linear layer. To develop the proposed hierarchical melanoma feature extractor, I choose ImageNet (Deng, Jia, et al. "Imagenet: 2009) pre-trained resnet34 (He, Kaiming, et al. 2019) as backbone.

Attention Module

The proposed Attention module only receives level 1 of the pyramid image and outputs an attention map which allows the model to focus on the region of interest in the melanoma input picture. I define the proposed Attention module as M_A: $I_1 \rightarrow A$. The attention map is then pixel-wise multiplied with the feature map which is the output of the M_F . The following equation explains how the pixel-wise multiplication is calculated.

Equation 1: Pixel-wise multiplication

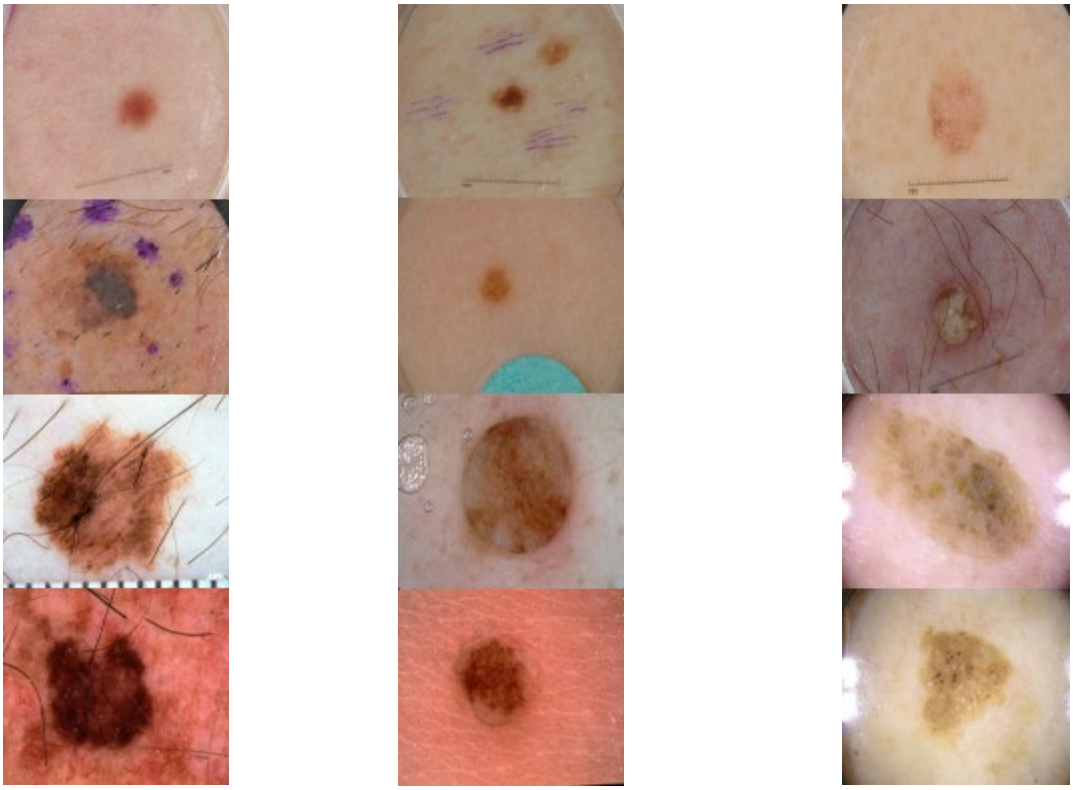
$$\boldsymbol{A * (F_1 + F_2 + F_3)}$$

Where the dot denotes the pixel-wise multiplication. The attention module enables the trained network to focus on the region of interest. The effectiveness of the proposed attention module is further studied in chapter 3.3. Finally, the result of the pixel-wise multiplication is the input of the linear layer. The linear layer then calculates the prediction vector into three levels, namely, (P_1, P_2, P_3) Where P_1 refers to the Seborrheic keratosis, P_2 refers to Melanoma, and P_3 refers to Nevus. I modified the Resnet38 architecture to design the proposed Attention Module. I remove the last linear layer and pooling layer.

Implementation Details

To implement the proposed system, I use Adam (Kingma, Diederik P. 2014) optimizer with beat=0.9 and beta=0.99. I set the batch size of twenty-four and the initial learning rate of 0.0001. I train the system for 100-epoch with learning rate decay at 40 and 80 epochs. For data augmentation, I proposed a novel skin color jitter augmentation technique to provide a wide range of skin color characteristics to the model during the training process. This skin color jittering helps the trained network with robustly inferencing for input variants. In chapter 3.3, I explain how the proposed augmentation technique contributes to the trained model.

Experimental Results



(a) Melanoma (b) Nevus (c) Seborrheic keratosis

Fig. 3. Example of each melanoma samples

Fig. 3. Shows the different types of melanoma samples that I used in this research. I use the Melanoma Detection Data (Codella, Noel, et al. 2019) set that is publicly available to train the proposed model. The samples in the dataset are labeled by specialists. The dataset is divided into three parts: testing, training, and validation. This dataset consists of about 2,750 sample images, with the test set having 600, training data with 2,000, and the validation with 150. Fig. 3-(a) is an example of melanoma, Fig. 3-(b) is a nevus, and Fig. 3-(c) shows seborrheic keratosis. The dataset includes a large variety of samples with different shapes and colors or melanoma which makes the classification particularly challenging. 70% of the samples are used to train the proposed model, 5% for validation, and about 20% to test the trained model.

Evaluation

For the evaluation metric, I measure the accuracy of the proposed method. I compare the proposed method with the state-of-the-art melanoma classification models that show comparable accuracy. I chose VGGNet (Simonyan, Karen, and Andrew Zisserman. 2014) and Densenet (Huang, Gao, et al. 2017), and Resnet (He, Kaiming, et al. 2019) which are publicly available and show good performance. Furthermore, I also conduct an ablation study to analyze the effectiveness of each proposed idea.

Table 1. Comparison with the state-of-the-art classification models

Method	Accuracy (%)
VGGNet (Simonyan, Karen, and Andrew Zisserman. 2014)	75.2
Densenet (Huang, Gao, et al. 2017)	77.9
Resnet (He, Kaiming, et al. 2016)	78.3
Ours	82.7

Table 1. shows the comparison results of the accuracy with the previous state-of-the-art methods. Our method has shown better performance compared to all the previous state-of-the-art methods. VGGNet (Simonyan, Karen, and Andrew Zisserman. 2014), as shown in table 1, yields an accuracy of 75.2%. The proposed method achieves 7.5% greater in accuracy compared to the first comparison model. The second method Densenet (Huang, Gao, et al. 2017) has an accuracy of 77.9%. The proposed methods outperform the second method with the accuracy gap of 4.8%. Finally, Resnet (He, Kaiming, et al. 2019), which shows the best performance among the comparison models, produces accuracy of 78.3%. The proposed method surpasses the third comparison model by accuracy of 4.4%. In conclusion, our proposed method outperforms all the previous state-of-the-art methods while showing a remarkable performance boost. These comparison results clearly show that the proposed architecture modification and data augmentation technique help with inferencing test images that have the shape and color variety in melanoma regions. The detailed analysis of each proposed idea is studied in chapter 3.3.

Ablation Experiments

Table 2. Effectiveness of each proposed idea

Model	Accuracy (%)
W/o hierarchical feature extractor	80.1
W/o attention module	80.4
W/o skin color augmentation	81.9
Full model	82.7

In this chapter, I conduct the ablation study to test and examine the effectiveness of each proposed idea. I first train the full model using all the proposed methods which are the hierarchical network module, attention module, and skin color jitter augmentation. I then compare the accuracy of the full model with three different ablation models. The ablation models are ordered as follows: One trained without the hierarchical network module, the second without the attention module, and finally without the use of skin color jitter augmentation.

As seen in table 2, the first ablation model without the use of a hierarchical feature extractor, dropped the accuracy by 2.6%. The performance gap is large, and I can safely assume that the hierarchical feature extractor plays a significant role in improving the accuracy of the model. The reason why a hierarchical feature extractor plays an important role is that the model divides the input image into three different resolution levels: high, medium, and low. The different resolutions of the input image allow the model to extract more rich and precise information about melanoma.

The second ablation model yields an accuracy of 80.4%. With a drop of 2.3% in the overall accuracy. Thus, the attention module is correlated to the increased effectiveness of our proposed model. As the attention module enforces the trained network to focus on the region of interest where melanoma might be located. This allows the model to focus on the area of the melanoma instead of the skin cells that are not a part of the melanoma.

The final ablation model, which is without skin color jitter augmentation, yields an accuracy of 81.9%. This is 0.8% lower than our proposed method. Here, I can also presume that the data augmentation has a role in increasing the performance of our model. This domain-specific data augmentation allows the trained model to perform more robustly against various types of skin color by preventing the model biased towards skin variations.

Conclusion

In this research, I proposed a melanoma skin cancer classification system. The proposed system is composed of the hierarchical feature extractor, the attention module, and the linear layer. I use three levels of pyramid images as input to extract more rich and accurate feature maps. The proposed hierarchical feature extractor takes these pyramid images as input and produces hierarchical feature maps. The attention module takes the level 1 pyramid image as input and outputs the attention map. This attention map is then multiplied by the hierarchical feature maps. Then, the Linear layer takes the aggregated feature map and predicts the class of the input melanoma image.

I also proposed a novel data augmentation to allow the trained model to see a wide range of color changes during the training process. Throughout the experiments, I have shown that the proposed method outperforms the existing state-of-the-art methods. I also conducted the ablation study to examine how each proposed idea affects the accuracy of the

trained model. For the dataset, I used the melanoma detection dataset which is publicly available online on Kaggle. For future work, I plan to make this application possible in real environments.

References

Argenziano, G., Cerroni, L., Zalaudek, I., Staibano, S., Hofmann-Wellenhof, R., Arpaia, N., ... & Kittler, H. (2012). Accuracy in melanoma detection: a 10-year multicenter survey. *Journal of the American Academy of Dermatology*, *67*(1), 54-59.

Adegun, A. A., & Viriri, S. (2019). Deep learning-based system for automatic melanoma detection. *IEEE Access*, *8*, 7160-7172.

Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., ... & Halpern, A. (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.*

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556.*