Learning How to Shade for Skin Lesion Diagnosis

Aaron Liu
Interdisciplinary Data Science Program
Duke University
Durham, NC
tienyu.liu@duke.edu

Christy Hu
Interdisciplinary Data Science Program
Duke University
Durham, NC
die.hu@duke.edu

Abstract

With skin cancer being one of the most common cancer in the US society, the number of patients who are diagnosed with skin cancer has been rapidly increased over the past decade. To prevent disease from advancing to later stages and thereby increasing the survival rate, there is demand for early detection with high accuracy, which sometimes can be difficult to achieve during the regular diagnosis process. Hence, in recent years the research community has geared efforts towards developing deep learning solutions to help with lesion diagnosis and benefit patients for quick and better results. Our results show that by carefully incorporating physical layer with illumination pattern extractor into convolutional neural networks can yield competitive classification accuracy while providing more explainable patterns to potentially help us design better physical devices and digital process for skin lesion diagnosis.

1 Instruction

Skin cancer is the most prevalent cancer in the United States[1][2]. Statistics suggest that approximately 9,500 people in the U.S. are diagnosed with skin cancer every day, and the number of new invasive melanoma patients diagnosed globally has increased by 47% over the last decade[3][4]. For patients whose melanoma have been diagnosed early, the average five-year survival rate is about 99 percent. However, when the infection spreads to lymph nodes, the survival rate slips to 65 percent and down to 25 percent if the disease metastases to remote organs[3]. It has been observed that[5] typically a professional dermatologist follows a sequence of steps when conducting the diagnosis, including a visual examination of suspicious lesion area, followed by dermoscopy and then biopsy. This whole process is time-consuming, and the disease may progress to later stages. In addition, accurate diagnosis is often dependent on the professional experience of dermatologists. Childhood melanoma is frequently postponed due to misdiagnosis of pigmented lesions, which happens up to 40% of the time.[6] All of the aforementioned problems call for early detection of skin cancer with improved accuracy.

To advocate for academic research in machine learning solutions to benefit the diagnosis of pigmented skin lesions, Tschandl released the HAM10000 ("Human Against Machine with 10000 training images") dataset[7], which contains a representative list of 7 major diagnostic types of pigmented lesions ranging from noncancerous or slow-growing skin cancer to malignant skin cancer such as melanoma. Previous research in the field of skin cancer classification has proven that deep learning solutions can help boost the accuracy to match or even exceed expert's performance. The goal of this study is to investigate how performance changes by effectively combining a simulated optical process and a deep convolutional neural network(CNN) with optimizable physical parameters in our imaging system. Our results show that this approach can yield competitive classification accuracy while providing more explainable patterns to potentially help us design better physical devices and digital process for skin lesion diagnosis.

2 Related work

There has been a lot of research effort to advance the field of skin cancer detection based on image analysis using machine learning and computer vision techniques. Researchers have sought to improve the accuracy of diagnosis by adopting different classification algorithms and techniques in all these attempts. Some of the most commonly used classifiers include support vector machines, Ada Boost, and decision tree algorithms. However, when convolutional neural network (CNN) took image classification task to new bounds, over the years it has been found that deep learning solutions can achieve performance on par with professional dermatologists when it comes to classifying skin cancer without having to perform feature extraction.

In 2017, Esteva et al.[8] made a breakthrough on skin cancer classification by using a pre-trained GoogleNet Inception v3 CNN model to classify 129,450 clinical skin cancer images comprising of 2,032 different diseases. In 2016, Yu et al.[9] developed a CNN with more than 50 layers using the ISBI 2016 challenge dataset and reported the best accuracy of 85.5% for classifying malignant melanoma. In 2018, Haenssle et al.[10] utilized a deep CNN to tackle dermoscopy image classification problem with binary class. In the same year, Dorj et al. [11] developed a multiclass classification solution by using ECOC SVM with pre-trained AlexNet Deep Learning CNN to classify 4 diagnostic categories of clinical skin cancer images, with an average reported accuracy of 95.1%. Han et al.[12] also attempted to classify clinical image of 12 skin diseases using a deep convolutional neural network.

Although current research in deep learning have achieved impressive results, the efforts were generally geared towards either optimizing the network architecture or algorithms at the digital processing stage. In this study, instead of aiming to push for maximum performance, we are more interested in learning how optimization of the physical parameter of the imaging system could impact our classification accuracy by simulating the physical world environment for capturing the skin lesion images.

3 Methods

The question of interest is to investigate the impact of incorporating a "physical" layer that simulates image-blurring filter that happens in the physical world, and whether adding an extra layer that performs color filtering or illumination pattern as digital process in a neural network can be used to contribute to the classification performance for lesion diagnosis. Regarding dataset used in this research project, all experiments were performed on the *HAM10000*'s training set since it provides a large volume and diverse collection of multi-source dermatoscopic images of common pigmented skin lesions for machine learning purpose [7]. This dataset consists 10015 dermatoscopic RGB images with image width 600 and height 450 pixels, and contains seven disease states: Pigmented Bowen's (*akiec*), Basal Cell Carcinoma (*bcc*), Pigmented Benign Keratoses (*bkl*), Dermatofibroma (*df*), Melanoma (*mel*), Nevus (*nv*), and Vascular (*vasc*). Sample dermatoscopic images are shown in Fig 1. In order to measure the generalization performance, the dataset is randomly partitioned into training (70%) and validation (30%) sets.

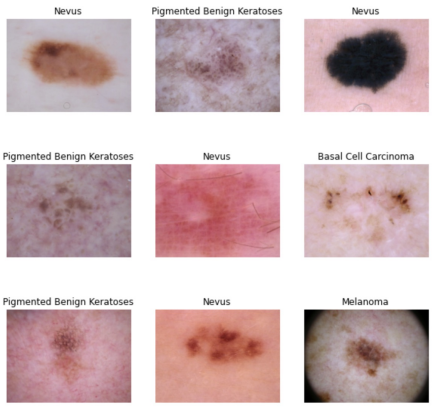


Figure 1: Sample dermatoscopic images

3.1 Physical layer

Although there are works that have been studied the impact of using Gaussian blur, we argue that such process may not be naturally observed in physical world. Therefore, instead of using Gaussian kernel, we pre-trained our blur kernel with a neural network on *HAM10000* dataset. The network architecture is formed by a sequence of layers: a re-scaling layer that brings unsigned integer to the range from 0 to 1 across RGB channels, 2D depthwise separable convolutional layer with square kernel size of 30 to simulate a blur kernel, a bi-linear resizing layer that down-samples the feature maps to size 32 by 32, 2D convolutional layer with 8 filters and kernel size of 3, ReLu activation, Batch normalization, Max-pooling, densely connected layer with 128 units, ReLU, densely connected layer with 7 units, and the output layer of using Sigmoid function. This network was trained using cross-entropy loss for 5 epochs with the ranger optimizer, reaching validation loss of 61.52%.

For the layer to provide color filtering, we use a 2D depthwise separable convolutional layer with kernel size 1. This operation can be treated as re-weighting the normalized input images along their RGB channels. For the layer of illumination pattern extractor, we use a matrix with size 450 by 600 that is repeated 3 times to match the number of input channels.

3.2 Convolutional neural network

3.2.1 Decoder block

The decoder block in this research project is formed by a convolutional neural network with a sequence of 3 mini-blocks of 2D convolutional layer with k filters and kernel size of 3, ReLu activation, Batch normalization, Max-pooling, where k is in the list of 8, 16 and 32. After these mini-blocks, we have densely connected layer with 128 units, ReLU, densely connected layer with 7 units, and the output layer of using Sigmoid function.

3.2.2 Ranger Optimizer

The ranger optimizer developed by Wright (`https://github.com/lessw2020/Ranger-Deep-Learning-Optimizer`) combines two very new developments, RAdam (Rectified Adam) [13] and LookAhead in one optimizer. RAdam is said to provide the best foundation for the optimizer to start training since it effectively provides an automatic warm-up to ensure that a solid training start can be obtained. LookAhead [14] on the other hand is believed to provide a breakthrough for robust and stable exploration throughout the training process. Therefore, ranger is expected to achieve the two kinds of improved best results by effectively combining these two.

In total, we are interested in studying the impact with 5 models (see below). For the training configuration, these models are trained for 10 epochs with batch size 32 using cross-entropy loss, and optimized by the ranger.

- **Baseline**: non-trainable blur kernel + decoder block
- **RGB filtering**: non-trainable blur kernel + color filtering + decoder block
- **RGB filtering***: trainable blur kernel + color filtering + decoder block
- **Illumination**: non-trainable blur kernel + illumination pattern extractor + decoder block
- **Illumination***: trainable blur kernel + illumination pattern extractor + decoder block

4 Results

The results show that the baseline model yields the highest accuracy of 75.33% (Table 1); meanwhile, we notice that both the models of RGB filtering and Illumination with trainable kernel provide competitive performance compared to those of with non-trainable kernel with accuracies of 74.20% and 73.13%. As we can see in Table 2, regarding loss, RGB filtering with trainable kernel yield the best loss of 0.7105, which outperforms the rest of the models, and RGB filtering and Illumination with non-trainable kernel again are ranked the last.

Based on the histories of validation accuracy and loss, the model of Illumination with trainable kernel seems to be able to provide competitive results in about 2 epochs. This implies that the pre-trained

blur kernel might be able to provide information to benefit generalization performance. In the history of accuracy, it seems that there is room for improvement in the baseline model as it was increasing while reaching the end of the training process.

Table 1: Best validation accuracy (%) in training history

Baseline	RGB filtering	RGB filtering*	Illumination	Illumination*
75.33	72.94	74.20	72.64	73.13

* Trainable blur kernel

Table 2: Best validation losses in training history

Baseline	RGB filtering	RGB filtering*	Illumination	Illumination*
0.7330	0.7626	0.7105	0.7713	0.7424

* Trainable blur kernel

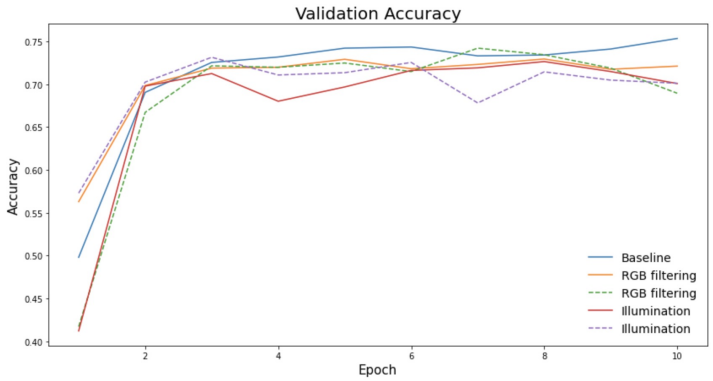


Figure 2: Validation accuracy under different experimental settings (dashed lines: trainable blur kernel)

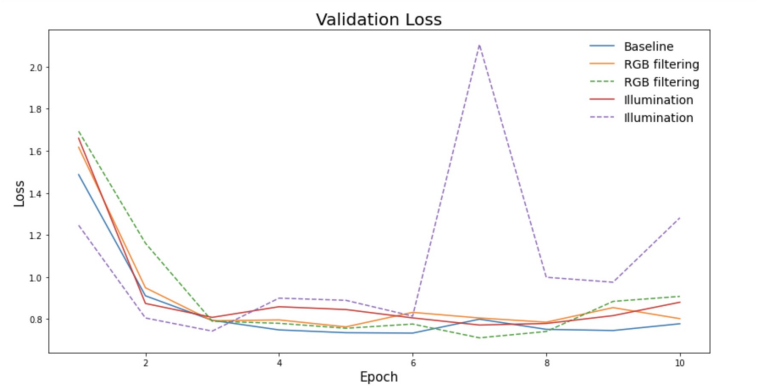
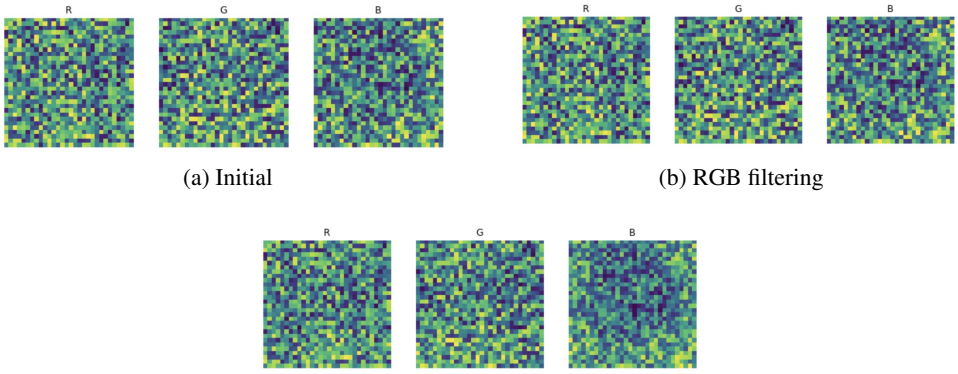


Figure 3: Validation loss under different experimental settings (dashed lines: trainable blur kernel)

5 Discussion

Even though the results show that the baseline model performs the best under our experimental settings. We argue that the newly optimized blur kernels in the models of RGB filtering and Illumination are providing explainable results that may help professionals to develop devices to behave like these kernels for more efficient diagnosis. In Fig 4, we can see that these newly optimized kernels along B channel have clearer circular shape compared to the initial kernel. This also implies that edge in B channel may provide more information in order to classify skin lesion. Regarding illumination pattern, in Fig 5, we think that both the optimized patterns with trainable and non-trainable blur kernel can be used to provide guidelines to help us design shading strategy to develop specialized convolutional neural networks, and give instruction to form better illuminating condition for more accurate skin lesion diagnosis.

For future work, firstly, as it has been shown that training longer may close the generalization gap [15], we would try to increase the number of training epoch to accommodate the larger hypothesis spaces in the models of RGB filtering and Illumination. Secondly, the pre-trained blur kernel acting as feature extractor may potentially hinder classification performance based on how it was produced; therefore, different network architecture designs may be considered to further explore the potential of incorporating physical layers in neural networks for skin lesion diagnosis.

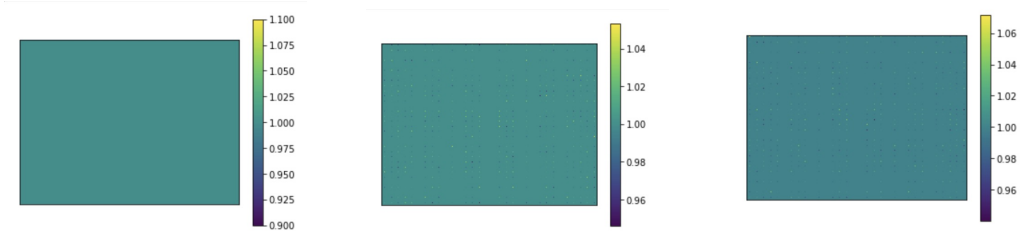


(a) Initial

(b) RGB filtering

(c) Illumination pattern

Figure 4: Blur kernels



(a) Initial

(b) Non-trainable kernel

(c) Trainable kernel

Figure 5: Illumination patterns

References

[1] G. Guy, C. Thomas, T. Thompson, M. Watson, G. Massetti, and L. Richardson, "Vital signs: Melanoma incidence and mortality trends and projections - united states, 1982-2030," *MMWR. Morbidity and mortality weekly report*, vol. 64, pp. 591–6, 06 2015.

[2] G. Guy, S. Machlin, D. D. Ekwueme, and R. Yabroff, "Prevalence and costs of skin cancer treatment in the u.s., 20022006 and 20072011," *American Journal of Preventive Medicine*, vol. 48, 11 2014.

[3] American Cancer Society, "Cancer facts and figures 2020." [Online]. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf

[4] ——, "Cancer facts and figures 2010." [Online]. Available: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2010.html

[5] M. A. Kadampur and S. Al Riyaee, "Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images," *Informatics in Medicine Unlocked*, vol. 18, p. 100282, 2020.

[6] A. Ferrari, A. Bono, M. Baldi, P. Collini, M. Casanova, E. Pennacchioli, M. Terenziani, I. Marcon, M. Santinami, and C. Bartoli, "Does melanoma behave differently in younger children than in adults? a retrospective study of 33 cases of childhood melanoma from a single institution," *Pediatrics*, vol. 115, no. 3, pp. 649–654, 2005.

[7] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, p. 180161, 2018. [Online]. Available: https://doi.org/10.1038/sdata.2018.161

[8] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017. [Online]. Available: https://doi.org/10.1038/nature21056

[9] L. Yu, H. Chen, Q. Dou, J. Qin, and P. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, 2017.

[10] H. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. Hassen, L. Thomas, A. Enk, and L. Uhlmann, "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of oncology : official journal of the European Society for Medical Oncology*, vol. 29, 05 2018.

[11] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9909–9924, 2018. [Online]. Available: https://doi.org/10.1007/s11042-018-5714-1

[12] S. Han, M. Kim, W. Lim, G. Park, I. Park, and S. Chang, "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *Journal of Investigative Dermatology*, vol. 138, 02 2018.

[13] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," 2020.

[14] M. R. Zhang, J. Lucas, G. Hinton, and J. Ba, "Lookahead optimizer: k steps forward, 1 step back," 2019.

[15] E. Hoffer, I. Hubara, and D. Soudry, "Train longer, generalize better: closing the generalization gap in large batch training of neural networks," 2018.