Melanoma Detection Using Deep Learning

Ameya Angal
Indiana University, Computer Science Department
Bloomington, Indiana
aangal@iu.edu

Akshay Naik
Indiana University, Computer Science Department
Bloomington, Indiana
aunaik@iu.edu

Ninaad Joshi
Indiana University, Computer Science Department
Bloomington, Indiana
ninjoshi@iu.edu

Siddharth Pathak
Indiana University, Computer Science Department
Bloomington, Indiana
sidpath@iu.edu

ABSTRACT

Melanoma is the most dangerous type of skin cancer, affecting 91,270 adults in 2016 with an increase of 3% per year from 2004 to 2014. [1] Detecting Melanoma is challenging, thus, Dermatoscopy is used to obtain high quality images which helps in detecting cancer in its early stages. Until now, it was the Dermatologists who were responsible to detect if the skin infection is cancerous or benign. This puts Dermatologists in high demand and they need to do a lot of work even for cases which can be benign. Thus, we intend to provide a good way to filter out the benign cases, which would make the work of Dermatologists much more efficient. Advances in Deep Learning techniques and Image Processing make it possible to detect Melanoma with high accuracy. Our proposed model will be able to detect whether a skin patch is malignant or benign from a given input image.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision**; **Visual inspection**; **Transfer learning**; **Neural networks**; • **Applied computing** → Imaging;

KEYWORDS

VGG-19 Neural Network, ResNet Neural Network, Melanoma Detection

1 INTRODUCTION

Melanoma is one of the deadliest types on skin cancer which affects more than 90,000 people per year. More than 10,000 people are killed because of Melanoma every year in the US. Thus, early detection and treatment is highly essential to eradicate the cancer. Dermatologists are required for the detection of the cancer, which puts them in high demand. However, if there could be a way to tell a normal skin infection from a malignant infection, it would be beneficial for the Dermatologists to improve their efficiency and they would be able to focus more on the real cancerous cases which are more important.

Dermatoscopy is a technique for capturing high resolution magnified images of the skin patches, which is used quite extensively by Dermatologists. The technique involves focusing and magnification on the skin patch, the use of a polarized light source, and a liquid medium between the instrument and the skin. This technique manages to filter majority of skin reflections thus providing highly detailed images. This is an effective and non invasive technique to filter out the benign cases from the malignant ones and has been proven to be quite accurate in terms of detecting the malignant skin lesions.

However, highly skilled and specialized personnel is required to operate the machines. Thus, there is a need to provide a novel solution which could detect the presence of malignancy from input images. Also, this solution needs to be robust to avoid false negative classifications which would be dangerous.

We propose a model using Deep Neural Net for detection of Melanoma from input images, which is reasonably accurate and would be helpful for doctors as well as patients for early diagnosis of skin infections and would detect if the given input images has Melanoma or not.

2 LITERATURE REVIEW

As mentioned previously, Melanoma is a deadly type of cancer and early diagnosis and detection are key to reducing the mortality due to Melanoma. Thus, quite a few number of techniques have been proposed in different papers using different techniques like Independent Histogram Pursuit[2], automatic segmentation using optimal color channels and hybrid thresholding technique[3], as well as deep learning techniques using Fully Convolutional Residual Network (FCRN)[4] for classifying melanoma skin lesions. The International Skin Imaging Collaboration, also known as ISIC, started a competition with different challenges for the competitors. They provided a large dataset to the participants which included labelled images of benign and malignant skin lesions. One of the ISIC challenges was to classify the images according to the labels which include different classes like benign, Melanoma, intermediate and unclassified, etc. For our problem, we decided to do binary classification of the data which would tell if the input image has melanoma or is benign.

Our approach was inspired from the need to improve the accuracy of detecting Melanoma which would at least be on par if not better than the accuracies of other approaches. Thus, we decided to use Deep Neural Networks for our approach.

Figure 1: Bening Image resized

Figure 2: Malignant Image resized

3 DATA DESCRIPTION

Data was collected from The International Skin Imaging Collaboration[1]. The dataset consisted of approximately 13000 images, along with their metadata about the age and sex of the corresponding image. The images are of very high resolution of about 2000 x 3000 size. The data is highly skewed with 12668 images for Benign images and around 1173 Malignant images. Also, the images don't contain the infected area/skin always at the center of the image, so we won't be able to directly crop out the center of the image directly.

To run the Neural Network efficiently, we have to reduce the size of images to make them more concise and small without losing the important features. The first approach we tried to reduce the size was to detect the infected section in each photo and then crop that section alone. For this approach, first, we tried to dilate each image and then find contours and select only the contour which contains the infected section. But this approach was not suitable as each image differed on the location and size of the infected region. To crop the image we also tried Blob detection technique to detect the infected region, but this required changing the parameters for the blob detector and not a single set of parameters satisfied all the images. Due to these limitations, we did not go forward with the approach of cropping the infected region.

Then to make the images smaller we used OpenCV resize function with interpolation technique as Inter-Area. This technique is preferred to make the images smaller in size as it gives moire-free results. It resamples the images using pixel to area relation. Using this technique we were able to resize the images without losing image features. As VGG 19 network is optimized for images with 224*224 size, we resized all the images to 224*224 size.

Initial dataset was very skewed with around 12,000 images of Benign class and around 1000 images of Malignant class. As neural networks work poorly when the dataset is skewed and will predict the class having a large number of examples while training. So to make the dataset balance we have to augment the Malignant images to increase the number of input samples. For Augmentation, we tried to rotate each image by 90, 180 and 270 degrees. Then we flipped each of the generated images horizontally and vertically. As a result of this augmentation, we got around 8000 images for Benign class. So the final ratio for Malignant and Benign images in the final dataset was around 3:2.

From this final dataset, we used 95% for training the neural network and tested on remaining 5% data.

4 IMPLEMENTATION

We have used Keras framework to train the models and utilized transfer learning methodology by considering few of the well-known network architectures pre-trained on ImageNet dataset. Transfer learning is a design methodology wherein we use knowledge gained from training a model on one task for solving different tasks.

Once the pre-processing was done, we divided first the data into train and test with 95:5 split and then divided the training data further into training and validation data with 90:10 split. Initially started by training an AlexNet architecture on International Skin Imaging Collaboration (ISIC) dataset. AlexNet is an 8 layered deep Convolution Neural network with five convolution layers and three fully connected layers. After few runs, we obtained decent performance, but we were not that satisfied with the results and started researching more abut different networks which are better then AlexNet. We came across VGG nets and decided to use VGG-19 which is modification of Alexnet which is based on the principle that increasing the depth of the network will significantly improve the performance of the network.

Along with VGG-19 we also trained ResNet-50 which is a slightly different architecture whose basic building block is ResNet block which will be explained later. It has been developed to overcome the drawback observed when the number of layers of plain networks were increased, it was observed that as the number of layers increased the performance of the network gets saturated and then it even starts degrading. The authors of Very Deep Convolutional Networks for Large-Scale Image Recognition [5] proposed a novel idea wherein instead of letting the stacked network learn unreferenced mapping we can make the network to learn residual mapping which is more easier to optimized as compared to the original unreferenced mapping. We have modified VGG-19 and ResNet-50 architecture as per our problem requirement.

4.1 VGG 19

VGG-19 architecture as shown in Figure 3 uses very small (3*3) convolution filters and a deep convolutional network with depth of 19 weight layers which include 16 convolution layers and three fully connected layers. The five max-pooling layers between the convolution layers carry out spatial pooling. We have replaced the original three fully connected layers with our modified layers, the

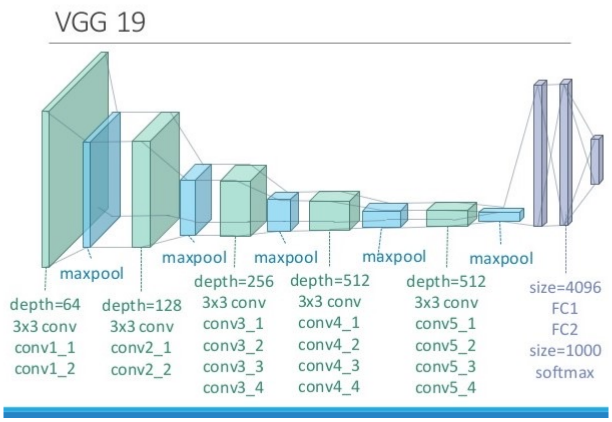

Figure 3: VGG-19 Neural Net model with 16 convolution layers and 3 fully connected layers
Figure Reprinted from: Applied Deep Learning, Mark Chang

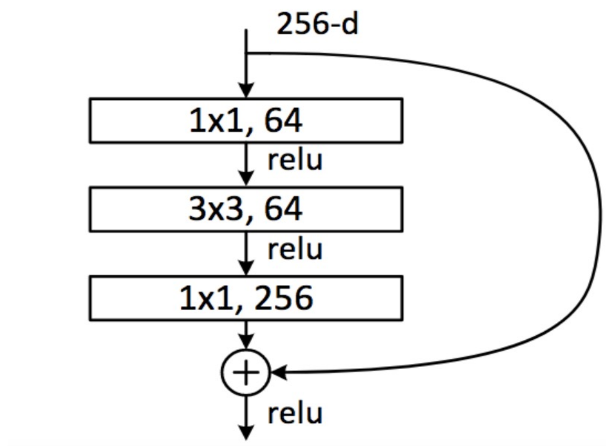

Figure 4: ResNet50 building block Figure reprinted from: Deep Residual Learning for Image Recognition, Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

first two layers having 1024 neurons and the last layer with a single neuron, as we are doing binary classification. We have used ReLU as activation function for first two fully connected layers and sigmoid for the output layer. Out of the 19 layers we are training last 14 layers. VGG with its increased depth in the network provides a significant improvement in the detection accuracy as compared to AlexNet (which had 83% Accuracy and 87% Recall).

4.2 ResNet-50

The main part in residual network is the residual block that stimulate the residual learning which help the network perform better. Residual block consist of two or more stacked layers with a shortcut connection helps the network to learn residual mapping. For ResNet-50, Figure 4, there are three stacked layers with a shortcut connection forming one residual block, as shown in above figure. We have replaced the output layer with a single neuron, as we are doing binary classification. We are training all the layers of ResNet-50 by initializing the weights of the network as pre-trained weights on ImageNet dataset. More details about residual network can be referred from [6]

5 RESULT

We trained both the models with different parameter values and displayed the top two results for both in Table 1 and plotted graph of accuracy and loss per epoch for the top performing parameters of both the models as shown in Figure 5 and 6. We observe that our VGG-19 model performs better with respect to both accuracy as well as recall. By observing the graphs, we can see that our ResNet-50 model overfits after 25 epochs, this might be the reason that our ResNet-50 model performs worse as compared to VGG-19. Theoretically, ResNet-50 model should have performed better than VGG-19 as its a deeper layer network, and also residual learning performs better than learning arbitrary function. We would like

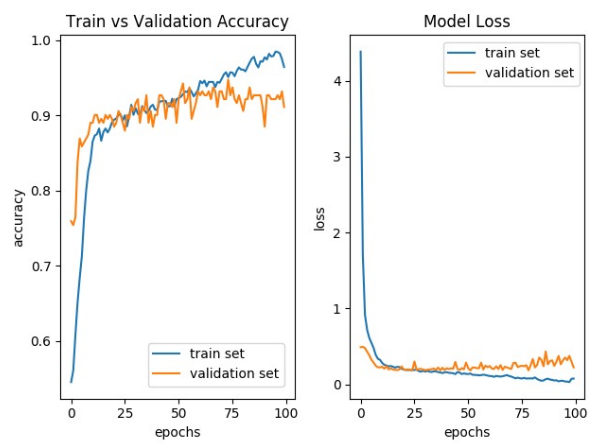

Figure 5: Accuracy and Model Loss for VGG-19

to investigate further about the ResNet-50 model to understand why it slightly over-fitted the train data. The result we obtained using VGG-19 is competitive and this model can be used to detect melanoma with a recall percentage of 94.12% which means 6% of time our model fails to detect melanoma.

6 EXPERIMENTATION

We started off by using AlexNet and its pre-trained weight on ImageNet dataset for classifying images. It gave us accuracy of 94.05% and recall value of 94.11% after fine tuning it. Further readings and research pointed us to use VGG19 - a newer and better version of AlexNet. VGG19 gave us better results than AlexNet, and thus we started fine tuning the hyper parameters of VGG19. We trained

Neural Net	Epoch	Number of Images	Learning Rate	Decay	Recall	Accuracy	Run Time(hrs)
VGG19	100 with Adam	2000	0.00001	1.00E-06	94.12	97.02	3
VGG19	100 with Adam	20000	0.00001	0	92.37	94.47	20
ResNet-50	100 with SGD	2000	0.0001	1.00E-06	89.47	93.07	4
ResNet-50	200 with SGD	16000	0.0001	1.00E-06	93.72	96.01	22

Table 1: Results

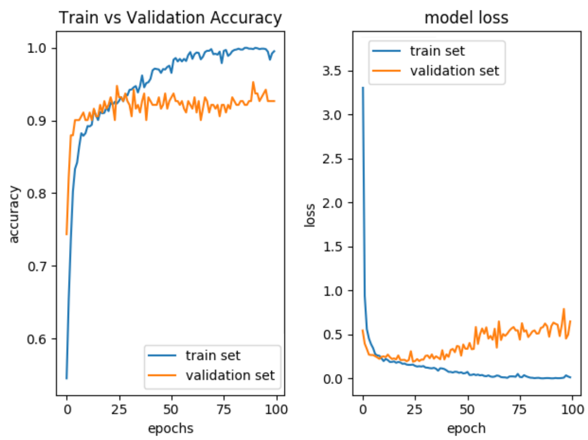

Figure 6: Accuracy and Model Loss for ResNet

VGG-19 with different values for the following parameters: Learning Rate, Decay Factor, Batch Size, Number of input images and number of layers trained.

For tuning the parameters, we started with less number of images [7], and gradually increased the number of images. For optimizing we used Adam optimizer with and without decay. We started observing how the loss is behaving per epoch ad saw that is was over shooting after every few epoches. To tackle this, we started tuning the decay factor and the training rate. We kept the batch-size to be 64, as further increasing it caused the GPUs to run out of memory. We have not trained the first five layers of VGG-19, and directly used the pre-trained weights. The intuition behind it being that the low level features (e.g. edges, contour) learned by the neural network will be similar for almost all the image related tasks. We have discussed the results obtained from VGG-19 in a later section.

We also used ResNet50. Initially we used pre-trained weights on ImageNet dataset without training any of the ResNet blocks and just by using bottleneck features as input to the output layer but the accuracy was less. We then started considering the number of convolution layers to be trained and consider it as a new hyper-parameter to be tuned, gradually increasing the number of layers to be trained starting from the last one layer. After varying the number of trained layers, we got the best accuracy when we trained all the 50 layers. We further tuned the model by using different optimizers, namely SGD and Adam. We got better results by using SGD, over Adam optimizer. Similar to what we had done for tuning VGG19, we used different values of learning rate, and decay. We had to limit the batch-size to 32, as higher batch-size caused the GPUs to run out of memory. We have discussed the results obtained from ResNet50 in a later section.

7 FUTURE WORK

Our current model does predict few false negative, we plan to further tune the model and decrease the number of false negatives, while also increasing the accuracy. Further tuning will also decrease the recall values.

Currently, we are classifying only if the cancer is of type Melanoma or not. We plan to gather more images, have our model classify different types of skin cancer. Benign types of infection can also be classified into different types if provided with the correct dataset. One of the potential applications of the trained model will be to develop a mobile app, which users can use to click photos directly from a smartphone, and instantly get a result if he/she needs to consultant a doctor.

8 CONCLUSIONS

We have presented a novel approach to detect Melanoma from high resolution images. We have used two Neural Net models namely VGG-19, and ResNet50 which can classify a given skin infection image into two classes – benign, and malignant, using the dataset consisting of approximately 13,000 images provided by ISIC. The best accuracy of VGG-19 model on the dataset was 97.02% with a recall value of 94.12%. While the best accuracy given by ResNet50 was 96.01% and best recall percentage of 93.72%. We believe that our implementation of VGG-19, and ResNet50 can be used for accurate detection of Melanoma. We plan to further tune the parameters to increase the recall percentage, and also decrease the number of false negatives.

REFERENCES

[1] American Cancer Society 's (ACS) publication, Cancer Facts Figures 2018

[2] D. D. GÃşmez, C. Butakoff, B. K. ErsbÃÿll, and W. Stoecker, *Independent histogram pursuit for segmentation of skin lesions*, IEEE transactions on bio-medical engineering, vol. 55, no. 1, pp. 157-161, 2008.

[3] R. Garnavi, M. Aldeen, M. E. Celebi, G. Varigos, and S. Finch, *Border detection in dermoscopy images using hybrid thresholding on optimized color channels* Computerized Medical Imaging Graphics, vol. 35, no. 2, pp. 105-115, 2011.

[4] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks IEEE Trans. Med. Imaging, vol. PP, no. 99, 2016.
[5] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*, 2014 arXiv:1409.1556.
[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep Residual Learning for Image Recognition, Dec. 2015.
[7] Discussed with Prof. Crandall to train on different size of dataset.