ISIC 2019 Skin Lesion Analysis Towards Melanoma Detection Challenge

Saamahn Mahjouri[1] and Raymond Phan[2]

[1] SharpestMinds, Minneapolis MN, USA
[2] SharpestMinds, San Francisco CA, USA
mahjouri.saamahn@gmail.com, rphan@ryerson.ca
https://www.sharpestminds.com/

Abstract. In this paper, we discuss our approach to solve the ISIC 2019 Skin Lesion Analysis Towards Melanoma Detection Challenge which consists of classifying eight types of skin lesions with an unknown class and with additional metadata to bootstrap the data. The provided dataset is also highly imbalanced and can plague many classification systems thus introducing heavy bias in finding the optimal classification model. Our approach is a novel as it combines methods to address imbalanced datasets and transfer learning using well-established deep learning models with incremental improvements by tuning the learning rate after groups of epochs. We also illustrate a workaround for the classification model to detect skin lesions that are of the unknown type which do not belong in any of the conventional labels.

Keywords: Image Analysis · Melanoma Detection · Skin Lesion · Deep Learning · Machine Learning · Image Classification

1 Introduction

Before we introduce the problem and how we address solving this problem, it should be noted that this manuscript only contains solving the first challenge, which uses just the images in the dataset. The second part of the challenge which includes metadata is not addressed as of yet. The manuscript will be modified accordingly to address incorporating the metadata when that task has been completed.

1.1 Overview of the Dataset

Table 1 lists each skin lesion type combined with their known acronyms in this challenge combined with their quantities. In total, there are 25,331 training images and with ground truth labels and 8,238 unlabelled test images. Each image is assumed to contain only one disease and is for only one patient.

Table 1. The ISIC 2019 Dataset

Skin Lesion Type	Acronym	Quantity
Melanoma	MEL	4522
Melanocytic nevus	NV	12875
Basal cell carcinoma	BCC	3323
Actinic keratosis	AK	867
Benign keratosis	BKL	2624
Dermatofibroma	DF	239
Vascular lesion	VASC	253
Squamous cell carcinoma	SCC	628
None of the others / Unknown	UNK	0

1.2 Data Preprocessing

Take note that there are no unknown images in the training set yet part of the challenge is to identify unknown cases in the test set. In addition, the metadata provided for the training set consists of the approximate age, the location of where the skin lesion was found and a unique identifier for each lesion of each patient. With the provided metadata, there are duplicate entries per image and thus patient meaning that it may be possible to have more than one class per patient. Due to the nature of the challenge where it is guaranteed that there is only one disease that exists per patient, we have taken the liberty of isolating out the images and thus patients that have only one identified lesion when consulting the metadata. Therefore, this decreases the amount of image data we are dealing which will thus make training easier.

1.3 Data Augmentation

By consulting Table 1, we can clearly see that the dataset is highly imbalanced. To properly ensure that the classification model we create is not biased towards examples that consist of a large proportion of the dataset, some form of data augmentation is required. After the removal of the duplicates in the data preprocessing step, due to the inevitable loss of data we decided to perform data augmentation so that the total number of images in each category approximately match the original number of images that belong to the Melanocytic nevus or NV class. Therefore, for the rest of the training images that belong to the other classes, we perform data augmentation on each class individually with the parameters shown in Table 2 up until each class reaches about 12900 images so that we now have a balanced dataset for creating the optimal classification model. As such, for each class to augment, we randomly select an image then randomly decide a value for each attribute in the table and apply the corresponding transformations to that image and save it. The convention of the table is such that W, H stands for the width and height an image respectively and the zoom factor is such that the image is randomly magnified between a factor of 0 to 0.1 and cropped to a desired size. The expected input size for ResNet50 are 224 $\times$ 224 images. However, we decided to increase the resolution to double in each dimension so the resized images are 448 $\times$ 448 instead to provide more information and higher resolution for the ResNet50 architecture to take advantage of. Therefore, the images are cropped to this size in addition to the images in the dataset being resized to this target size. There are additionally brightness and contrast operations to applied to the image. Assuming the image has normalized RGB values from $[0, 1]$, we either add a random offset to all colour channels and/or scale all of the values by a random value, both with probability $p = 0.75$.

Table 2. List of transformations performed on an image in data augmentation

Transformation	Range of Values
Rotation	$[0^o, 180^o]$
Horizontal Translation	$[0, 0.1W]$
Vertical Translation	$[0, 0.1H]$
Zoom Factor	$[0, 0.1]$
Horizontal Flip	[`true`, `false`]
Vertical Flip	[`true`, `false`]
Brightness	$[0.45, 0.55], p = 0.75$
Contrast	$[0.9, 1.11], p = 0.75$

2 Classification Model

Deep Learning for Computer Vision applications has revolutionized the way we approach problems in this space, and in particular Image Classification. This line of thinking can also be applied to this challenge where the goal is

to identify what skin lesion each image belongs to. We can thus treat this as a machine learning one but instead of performing feature engineering, we can use established deep learning architectures to help do the job for us where training over many images helps derive the most optimal features so that classification can be performed with ease. Inspired by deep learning, we use the ResNet50 convolutional neural network [1] as the main classification model. This deep learning framework was trained on significantly different images, and so we employ transfer learning [4] to use the previous acquired knowledge about the images that the ResNet50 architecture has seen and bootstrap it with the skin lesion images that are seen in this challenge. The parameters learned from the ResNet50 architecture are a good starting point and should be modified to suit the images found in this challenge. Therefore, the snapshot of the original parameters from ResNet50 are used as a starting point, then are optimized using the images in the challenge's dataset.

We begin our transfer learning process by first splitting up the images into a training and validation split with an 99%/1% ratio. It should be noted that this is done on the augmented images found in the data augmentation stage. Due to the inherent difficulty in training very deep neural networks, we employ the work by Smith [2] where we train the network using standard Stochastic Gradient Descent [3] to combat choosing the most optimal learning rate. Smith's Cyclical Learning Rate methods have shown faster convergence for ResNet50 training than using a standard constant learning rate. As such, we first employ Smith's `1cycle` policy by setting the number of epochs in one cycle to be 15 while we vary the learning rate between 1×10^{-6} to 1×10^{-3}. The selected batch size for optimization was 64. `1cycle` allows us to simultaneously optimize the parameters in the ResNet50 model with the varying learning rate to promote faster convergence. To further account for the unknown (UNK) label in the test dataset, we simply assume that every label in the test dataset has a ground truth label of UNK which also forms part of the validation set. The optimization of the parameters is done solely on the training set while the validation set created from the data augmentation in addition to the test set serve only to provide accuracy and examining how the validation loss behaves over each epoch.

3 Results

Figures 1 and 2 illustrate the loss trend over each batch and the confusion matrix derived from the combined validation dataset. The loss trend between training and validation seem to track each other which does not lend to overfitting. What is interesting to see is the UNK label - due to the absence of the actual ground truth labels for the test dataset, the row illustrating the UNK label classifications theoretically gives us a rough idea of what the actual labels should be. In addition, the training accuracy is 89.75% for the training data while the validation data was 90.77%. In order to better provide predictions that are agreeable with pathologists, upon inference of the test dataset we replace the Softmax output layer with a Sigmoid layer instead which provides a fuzziness measure for each class which thus serves as our final predictions to be submitted for the challenge.

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. IEEE Conf. on Computer Vis. and Pattern Rec. (CVPR) (2016), `https://arxiv.org/abs/1512.03385`
2. Smith, L.N.: Cyclical learning rates for training neural networks. arXiv (2017), `http://arxiv.org/abs/1506.01186`
3. Stanford Deep Learning Laboratory: Optimization - Stochastic Gradient Descent, `http://deeplearning.stanford.edu/tutorial/supervised/OptimizationStochasticGradientDescent/`, accessed on August 10, 2019
4. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C.: A survey on deep transfer learning. arXiv (2018), `http://arxiv.org/abs/1808.01974`

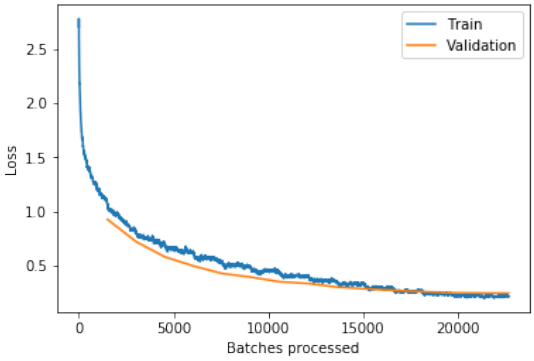

Fig. 1. Training and validation losses

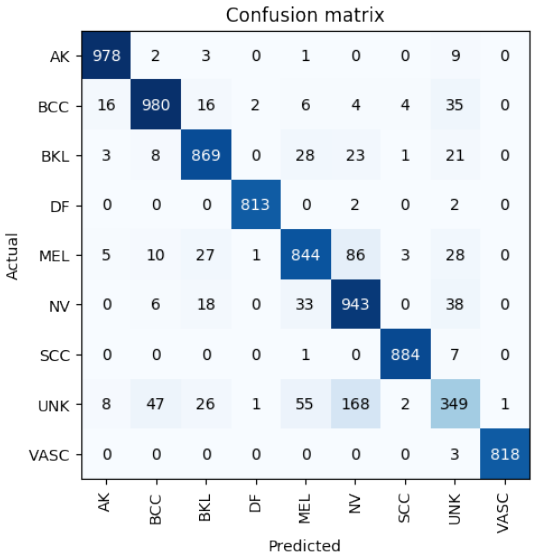

Fig. 2. Confusion matrix on the combined validation set