Report

November 28, 2018

1 Skin Cancer Classification Using Deep Learning

1.1 I. Introduction

In the United States, 5 million new cases of skin cancer are diagnosed every year.[1] Of these, melanoma which is the deadliest accounts for over 9000.[2] The diagnosis via visual inspection by patients and dermatologists is accurate only about 60% of the time.[3] Moreover, the shortage of dermatologists per capita has abetted the need for computer-aided methods to detect skin cancer.[4]

The International Skin Imaging Collaboration (ISIC) has aggregated a large amount of publicly accessible dermoscopy images labeled with ground truth data. The ISIC 2018 challenge [5] was divided into 3 tasks - Task1: Lesion Segmentation, Task 2: Lesion Attribute Detection and Task 3: Disease Classification. This report focuses on Task 3 i.e. classification of images into one of 7 possible classes.

1.2 II. Dataset

There are 10,015 images in the labeled training dataset. Some sample images from the dataset and their labels are shown in Fig. 1. The labels are in a CSV file in the form of one-hot vectors. There are no missing labels and all images are classified into one of 7 classes: * Melanoma * Melanocytic nevus * Basal cell carcinoma * Actinic keratosis / Bowen's disease (intraepithelial carcinoma) * Benign keratosis (solar lentigo / seborrheic keratosis / lichen planus-like keratosis) * Dermatofibroma * Vascular lesion

Fig. 1: Sample Images from Task 3 Dataset with Labels

Evaluation metric for this task is the multi-class accuracy (MCA) i.e. the average of precision for all classes.

$MCA = \sum^n \frac{P_i}{n}$

where P_i is the precision of class i

1.3 III. Architecture

Since this is a problem of image classification, a convolutional neural network (CNN) architecture would be most suitable. We used existing models such as VGG19[6], SqueezeNet[7], Resnet50[8], Inception[9] with weights pretrained on ImageNet[10][11]. We found ResNet50 to give the best results among these models.

Since ImageNet uses an input size of 224x224 and the images in the dataset are 450x600, the first step was to scale the images down to the required 224x224 size

1.4 IV. Improving Performance

Approximately 70% of the images belong to only one class (Melanocytic nevus). Hence, it is trivial to achieve around 70% accuracy by simply predicting all images to be of that class. That is obviously incorrect. In order to improve performance, we try several techniques such as data augmentation, oversampling low-frequency classes, weighted loss etc.

As expected a baseline model with 10% of the labeled dataset randomly kept aside as a validation set achieved only 62.74% MCA while the training MCA was 93.67%. We will attempt to improve this discrepancy in the performance of training and validation set by using some techniques as follows.

Fig. 2: Distribution of classes in the training set

1.4.1 A. Batch Normalization

Since we are using ReLU activoation in the fully connected layers and the final output is a softmax i.e. a number between 0 and 1, batch normalization [12] could help speed up training by scaling the output of the fully connected layers appropriately. The performance on the validation set improved slightly to 73.29% but the training MCA went down to 87.77%.

1.4.2 B. Data Augmentation - Mirroring

There is still a large difference between the validation and training MCA. Training the model on a larger dataset could help bridge this gap. We can double the dataset by simply taking mirror images [13] of the existing dataset while keeping the labels constant.

1.4.3 C. Weighted Loss

The model still predicts the dominating class more often than it should while ignoring lesser occuring classes. One way to fix this is to penalize the model for predicting the dominating class.[14] This can be done by multiplying the loss function by the frequency of classes. Thus, a new weighted loss function can be used to train the model.

1.4.4 D. Oversampling

The presence of classes Dermatofibroma and Vascular lesion is very low in the dataset (~0.01%). We can increase their occurence by taking random crops of the central part of the image so that the lesion still remains in the image.[15] We took 4 random crops of images belonging to these classes and also their mirror images while keeping the labels constant. These were then added to the dataset from which 90% of the data was randomly selected for training.

1.4.5 E. Color Constancy

1.5 V. Results

The results achieved from training using the techniques listed above are shown in Table 1.

Technique	Training MCA	Validation MCA
Baseline	93.67%	62.74%
Batch Normalization	87.77%	73.29%

Technique	Training MCA	Validation MCA
Data Augmentation - Mirroring	70%	70%
Weighted Loss	70%	70%
Oversampling	70%	70%
Color Constancy	70%	70%

1.6 VI. Conclusion and Discussion

1.7 References