University of Padua

Department of Computer Science

Skin cancer classification using Keras

Alberto Bezzon 1211016

Tommaso Carraro 1210937

The report resumes briefly the cognitive services project

Academic year 2018 - 2019

1 Introduction

Skin cancer is the most widespread cancer and one of the most dangerous because of the number of cases that not only exceeds the combined total of new cases for prostate cancer, breast cancer, lung cancer, and colorectal cancer, but also it increases from year to year. Malignant melanoma is a prevalent type of cancer that is especially deadly. It is well known that early detection and proper treatments for new malignant skin cancer cases are very important to ensure high survival rate. Indeed, with an appropriate treatment in an early stage, survival rates are very promising. Otherwise, the survival rate for melanoma decreases from 99% to 14% in more advanced stages.

The usually way to detect a melanoma is by inspecting the visual details of skin which has a low precision. Another way is dermoscopy, a non-invasive technique, that can capture a high resolution image of the skin which enables dermatologists to detect features which are invisible to the naked eye. This technique makes easier to diagnose melanoma, but it is time consuming and it is based on the skill of clinician that made dermoscopy. Moreover, because of the resemblance between malignant skin tumors and benign skin lesions in visual features, it is very hard for dermatologists to differentiate between them.

In recent years, deep learning, and specifically convolutional neural networks (CNNs), have reached very good performance in skin cancer classification tasks and have allowed computers to outperform dermatologists. For this reason, in our project we have tried to improve the accuracy of skin cancer detection using state-of-the-art CNN models and techniques. In particular, we have used these models to distinguish between seven common types of skin cancers that are included in the HAM10000, a recent and famous dataset made for this specific task. We obtained the best performance on the test set using data augmentation to train a simple CNN model built from scratch.

This document is organized as follows:

- Section 2 presents related works for "Skin cancer classification";
- Section 3 provides the description of dataset used in our experiments;
- Section 4 includes the approach used for solving the task;
- Section 5 presents the experiments we made;
- Section 6 contains the conclusion, the results of experiments and future works.

2 Related works

We have viewed different state-of-the-art papers. In [1] researchers used Googles Inception v3 CNN architecture pretrained on the 2014 ImageNet Challenge. They then removed the final classification layer from the network and retrained it with their dataset, fine-tuning the parameters across all layers. During training they resized each image to 299 299 pixels in order to made it compatible with the original dimensions of the Inception v3 network architecture. All layers of the network were fine-tuned using the same global learning rate and RMSProp optimizer. They performed their experiments on a 129.000 images dataset,

created from a combination of different datasets. They obtained 72.1% overall accuracy training their model on 757 classes. To create these training classes they used a taxonomy of skin disease and a partitioning algorithm that maps diseases into training classes. We have tried to obtain these data, but they were protected by the Stanford Hospital.

In [2] researchers studied the effectiveness and capability of different pre-trained state-of-the-art CNN architectures (DenseNet 201, ResNet 152, Inception v3, InceptionResNet v2). All the models they used were pre-trained on the 2014 ImageNet Challenge. They changed the classification part of these models with a custom classifier and they retrained them across all the layers using different hyperparameters depending on the specific network architecture. They trained these models on a dataset composed of 10.135 dermoscopy skin images composed by the combination of HAM10000 and PH2 datasets. The aim of their project was to compare the ability of deep learning with the performance of highly trained dermatologists. Overall, the mean results show that all deep learning models outperformed dermatologists (at least 11%). The best ROC AUC values for melanoma and basal cell carcinoma are 94.40% (ResNet 152) and 99.30% (DenseNet 201) versus 82.26% and 88.82% of dermatologists, respectively.

In [3] researchers used a CNN model built from scratch. The proposed model architecture consists on a sequence of alternating Conv2D and MaxPooling2D layers that form the core building blocks of modern CNNs. Then, they putted a batch normalization layer after each ReLu activation. They used this model to perform a binary classification task, in fact the first convolutional layer takes in 224 x 224 skin lesion images and the last dense layer contains a single unit with sigmoid activation in order to output the resulting classes (benign and malignant). They trained their model on the PHDB melanoma dataset, created by their own from a combination of open access datasets. They obtained an accuracy of the 86% on this dataset and regularization techniques such as dropout (0.5) and data augmentation techniques were heavily relied upon to combat the overfitting problem.

After reading these papers we decided firstly to try the power of transfer learning on our dataset. Given the big amount of data in HAM10000 and the high difference between the images in our dataset and the images included in the ImageNet Challenge, the best transfer learning strategy was to train the pre-trained model across all the layers. This technique gives us bad performance compare to the training of a simple CNN model built from scratch. So, we decided to use an approach that is more similar to the method explained in [3] than the other papers.

3 Dataset

On the Internet there are few open access datasets for skin cancer classification task, and most of them contain a collection of bad quality images that are not biopsy proven. We read a lot of papers on this task and in most of them researchers had to create a dataset from scratch for the unavailability of a complete dataset composed of quality and biopsy proven images on the web. In most cases they took best images from different sources and they built their own dataset. Recently, some researchers and dermatologists understood the importance of this task and decided to create a new dataset, called HAM10000, that we used on our project. This dataset contains 10015 dermatoscopic images that were collected over a period of 20 years from two different sites, the Department of Dermatology at the Medical University of Vienna, Austria, and the skin cancer practice of Cliff Rosendahl in Queensland,

Australia. It includes pigmented lesions from different populations. The Austrian image set consists of lesions of patients referred to a tertiary European referral center specialized for early detection of melanoma in high risk groups. The Australian image set includes lesions from patients of a primary care facility in a high skin cancer incidence area. Dermatoscopic images of both study sites were taken by different devices using polarized and non-polarized dermatoscopy. The set includes representative examples of pigmented skin lesions that are practically relevant. More than 95% of all lesion encountered during clinical practice will fall into one of the seven diagnostic categories contained in the dataset. In practice, the task of the clinician is to differentiate between malignant and benign lesions, but also to make specific diagnoses because different malignant lesions may be treated in a different way and timeframe. The number of images in the datasets does not correspond to the number of unique lesions, because experts also provide images of the same lesion taken at different magnifications or angles, or with different cameras. This should serve as a natural data-augmentation as it shows random transformations and visualizes both general and local features.

The seven different categories of skin lesions contained in the dataset are:

akiec Actinic Keratoses (Solar Keratoses) and Intraepithelial Carcinoma (Bowens disease) are common non-invasive, variants of squamous cell carcinoma that can be treated locally without surgery. There is agreement that these lesions may progress to invasive squamous cell carcinoma.

bcc Basal cell carcinoma is a common variant of epithelial skin cancer that rarely metastasizes but grows destructively if untreated.

bkl "Benign keratosis" is a generic class that includes seborrheic keratoses ("senile wart"), solar lentigo and lichen-planus like keratoses (LPLK), which corresponds to a seborrheic keratosis or a solar lentigo with inflammation and regression. From a dermatoscopic view, lichen planus-like keratoses are especially challenging because they can show morphologic features mimicking melanoma and are often biopsied or excised for diagnostic reasons.

df Dermatofibroma is a benign skin lesion regarded as either a benign proliferation or an inflammatory reaction to minimal trauma.

nv Melanocytic nevi are benign neoplasms of melanocytes and appear in a myriad of variants, which all are included in this dataset. The variants may differ significantly from a dermatoscopic point of view.

mel Melanoma is a malignant neoplasm derived from melanocytes that may appear in different variants. If excised in an early stage it can be cured by simple surgical excision.

vasc Vascular skin lesions in the dataset range from cherry angiomas to angiokeratomas and pyogenic granulomas. Hemorrhage is also included in this category.

In figure 1 is possible to see the difference between these types of skin lesions.

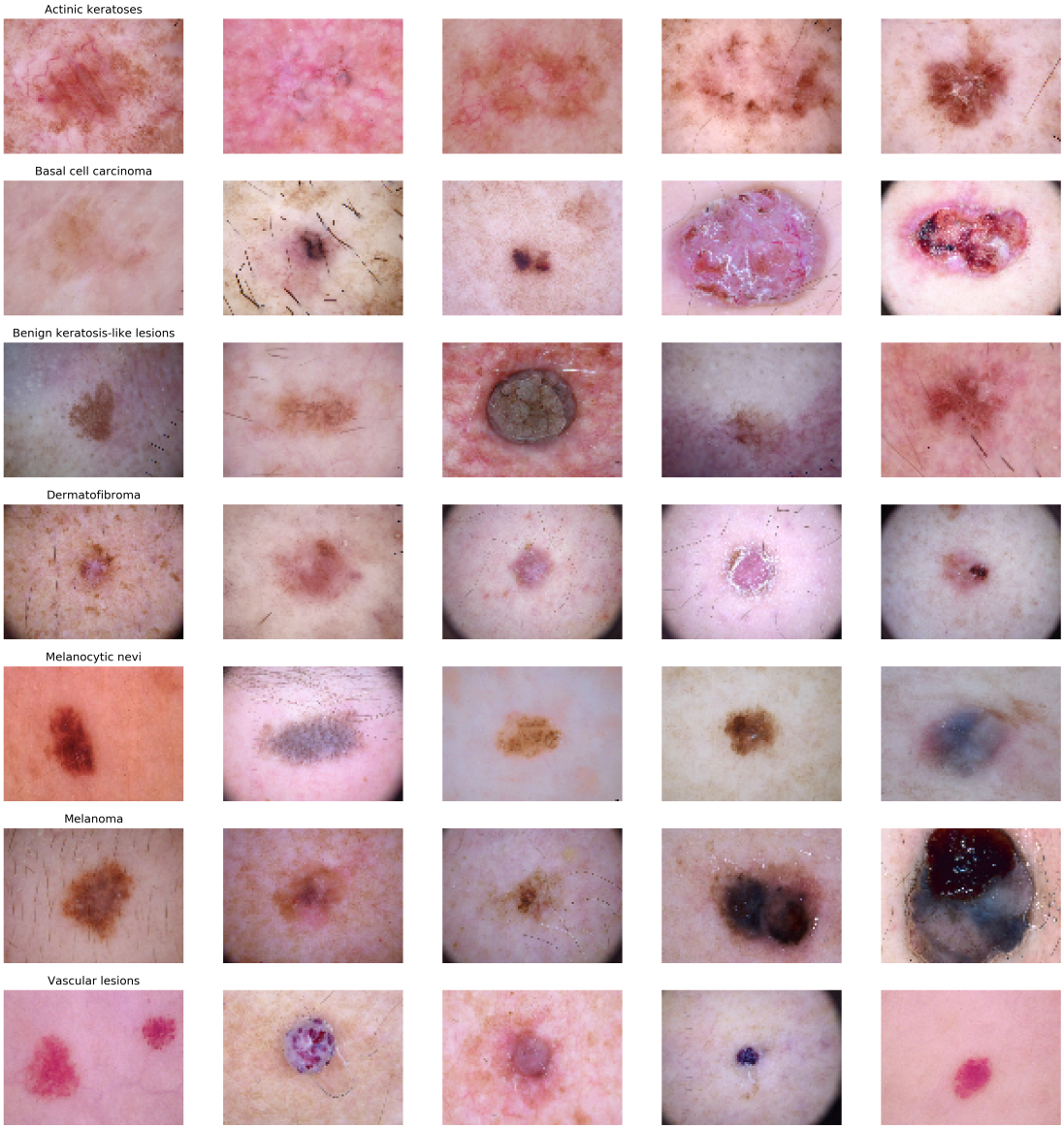

Figure 1: *Categories of skin lesions in HAM10000*

Unbalanced data is the biggest problem of HAM10000, in fact melanocytic nevi is the majority class with 67% of examples, instead dermatofibroma is the minority class with only 4% of examples. In figure 2 is possible to see the distribution of classes.

Unbalanced data is a common issue of skin cancer datasets and has been a challenge for this project. Initially we thought to find other images to oversample the dataset, but then we turned out that images in the internet are bad in quality and different compared to the images of the HAM10000, so we have decided to keep it at the original version. Data preprocessing in this dataset is limited to the normalization of images and the transformation of labels in one-hot encoded vectors to make the net capable to learn from them. We have decided

to split the dataset in 80% training set, 10% validation set and 10% test set. We kept this aggressive split because of the limited amount of data, in fact we want our model to learn as much as possible features from the dataset. Images have been resized to 100 x 75 before being supplied to the CNN model. We have decided this specific sizes because experiments showed that the model works better with small resolution images.

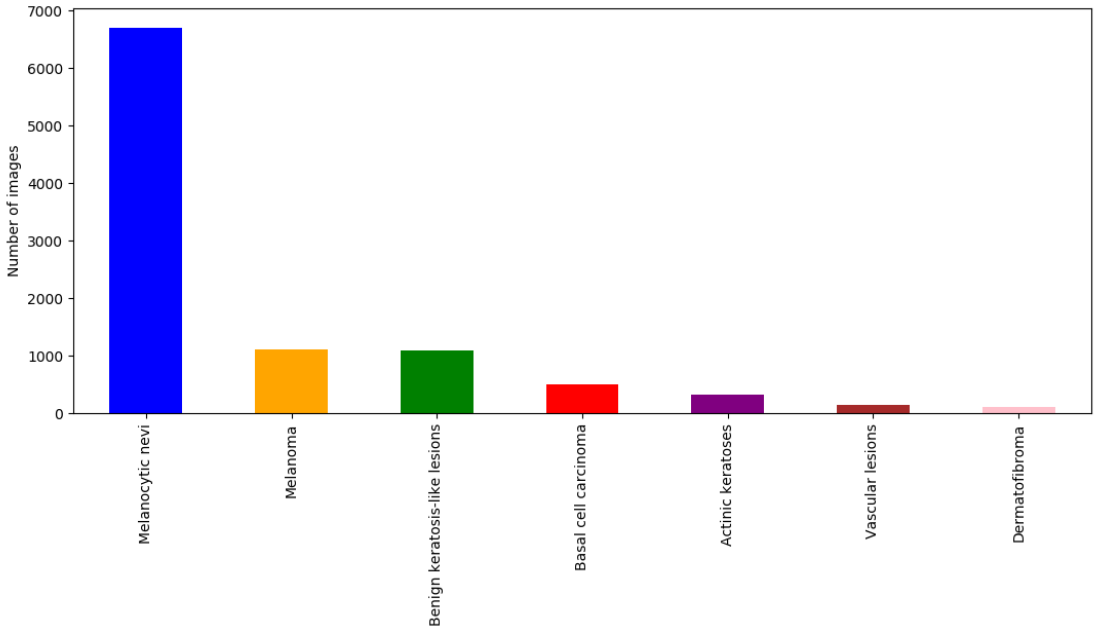

Figure 2: *Distribution of categories in HAM10000*

4 Method

5 Experiments

5.1 Experimental enviroments

Training a deep learning model that involves intensive compute tasks on large dataset can take days to run on a single CPU or a slow GPU. In our case, since HAM10000 dataset has 10015 images, it is unthinkable to perform the training of a convolutional neural network with a standard laptop. The solution turned to cloud computing. The choice fells on Google Cloud Platform because of the availability of free tier that consists in 300$ free credits that can be used in any GCP product.

We have tested our CNN models on a custom instance of Compute Engine. Our VMs configuration is presented in Table 1

Operating System	**CPU**	**Memory**	**Disk**	**GPU**	**Availability zone**
Ubuntu 18.04 LTS	8 core	52 GB	SSD / 100 GB	1x NVIDIA Tesla K80	europe-west1-b

Table 1: *Virtual machine configuration*

Our models have been implemented in keras using tensorflow as a backend. The code is available on our GitHub repository: https://github.com/albertobezzon/cognitiveservices

Bibliography

[1] Roberto A. Novoa Justin Ko Susan M. Swetter Helen M. Blau & Sebastian Thrun Andre Esteva Brett Kuprel. "Dermatologist-level classification of skin cancer with deep neural networks." In: (2017) (cit. on p. 1).

[2] Somayeh Karimijeshni Amirreza Rezvantalab Habib Safigholi. "Dermatologist Level Dermoscopy Skin Cancer Classification Using Different Deep Learning Convolutional Neural Networks Algorithms." In: (2018) (cit. on p. 2).

[3] Abhishek Verma Phillip Ly Doina Bein. "New Compact Deep Learning Model for Skin Cancer Recognition." In: (2018) (cit. on p. 2).