Skin Lesion Classification, IMA205 Challenge

Ramzi Charradi

April 2019

Contents

1 Introduction

A skin lesion is a part of the skin that has an abnormal growth or appearance compared to the skin around it. Skin lesions, such as moles or birthmarks, can degenerate and become melanoma, one of the deadliest skin cancer. Its incidence has been increasing during the last decades, especially in the areas mostly populated by white people.

The most effective treatment is an early detection followed by surgical excision. This is why several approaches for melanoma detection have been proposed in the last years (non-invasive computer-aided diagnosis (CAD)).

2 Problem statement

The goal of this work is to classify images of skin lesions as either benign or melanoma. However, this task is challenging due to the low contrast of skin lesions, the huge intraclass variation of melanomas, the high, degree of visual similarity between melanoma and non-melanoma lesions, and the existence of many artifacts in the image like hair or veins. Moreover, we have a very limited number of skin lesion images with their corresponding ground truth.

In order to detect melanoma images i used two approaches. First, I used Deep Learning since CNNs have defined the state of the art in image classification. Second, I extracted features such as the Asymmetry, the Border irregularity, the Color and the Dimension of the lesion (usually called the ABCD rule) and i used machine learning algorithms to classify the images.

3 Data

We used a data-set of 900 RGB images of skin lesions with their relative segmentation. Data has already been randomly split into a training-validation set (70%) and a test set (30%). We only have the classification (made by clinicians) of the training-validation set.

4 Deep Learning

4.1 Deep Residual Network

The first neural network i used to classify skin lesion images is a deep residual network proposed by Lequan Yu, Hao Chen, al. [1] in their work on skin lesion segmentation and classification for the ISBI 2016 Skin Lesion Analysis Towards Melanoma Detection Challenge.

Residual networks introduces extra skip connections to improve the information flow within the network, one motivation for skipping over layers is to

avoid the problem of vanishing gradients, by reusing activations from a previous layer until the adjacent layer learns its weights.

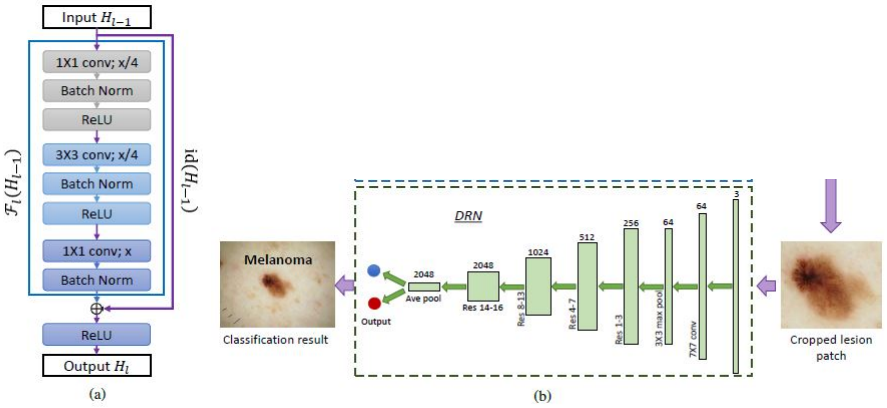

Figure 1: architecture of the residual network

Figure 1 (a) and (b) show the architecture of the DRN. This proposed residual network contains 4 residual blocks. Each residual block consists of two 1×1 convolutional layers, one 3×3 convolutional layer, three batch normalization layers and three ReLU layers, as shown in Fig. 1 (a). Besides these residual blocks, it also contains one 7×7 convolutional layer and one 3×3 max pooling layer.

This residual network showed excellent results in ISBI 2016 challenge whcih make from it a strong candidate for use in the IMA205 challenge.

4.2 VGG-16

VGG-16 is a deep convolutional network for object recognition developed and trained by Oxford's renowned Visual Geometry Group (VGG), which achieved very good performance on the ImageNet dataset. This architecture is a succession of convolution and max pooling layers.

In order to have the best results possible using VGG, I trained the network in two steps: First, i trained a shallow classifier on top of the VGG layers, That means that this classifier will use the features extracted with VGG to learn the weights, this helps speeding up the final training of VGG network.To do so, i dropped dense layers from VGG network, froze its weights and added a dense layer with a dropout rate of 0.5 and a final softmax layer for classification.

Second, after training this classifier on top of VGG features, i used all the potential of VGG and i trained the whole network on the skin lesion images. Now, the network is able to extract features related to skin lesion with the early convolutions of VGG and then provide a prediction for the class.

4.3 Training procedure

Given 700 training images, i used the first 600 images as a trainig set and the other 100 images for the validation. As preprocessing, images were cropped using the given segmentation in order to eliminate artifacts, then images were resized to the shape 224×224.
To increase robustness and reduce overfitting, i further utilized the strategy of data augmentation to enlarge the training dataset. The augmentation operators include rotation, translation and zoom.
Moreover, i modified the weights of each class in order to reduce the effect of imbalanced data, then i used stochastic gradient decent with a learning rate of 10^{-3} initially and reduced it by a factor of 10 every 20 epochs. I implemented the above networks with keras and i used colab GPU for computation.

4.4 Results

As expected, CNN performed well on the classification task either on training as on validation images.
For the evaluation i used accuracy and Matthews correlation coefficient :

$$AC = \frac{TP+TN}{TP+FP+TN+FN}$$

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{((TP+FP)(TP+FN)(TN+FP)(TN+FN))}}$$

As expected also, DRN took much more time to train the weights than VGG since it's a deep structure with a lot of convolution layers, for one epoch DRN took about one hour while VGG took 42 minutes to train. The table below resumes the performance of both architectures after 60 epochs of training and minimizing the binary cross-entropy loss function

Table 1: Performance of the networks on validation data

	loss	AC	MCC
VGG-16	0.49502	0.76	0.54
DRN	0.52102	0.69	0.44

5 Feature extraction

5.1 Features

Another approach consists in extracting features related to the shape, color and other specificities of lesions and then applying machine learning methods adapted with the problem. I extracted a dozen of features described in the work of Maciel Zorteaa, Thomas R. Schopfb, Kevin Thon et al. [2] and i added

some features using the work of Palash Sarkar on github `https://github.com/Tejas07PSK/Melanoma-Detection`. The features are :

- Asymmetry of shape (2 features)
- Asymmetry of color intensity (2 features)
- Colors, peripheral versus central (6 features)
- Features related to border, Moments, energy, entropy, contrast, color spaces statistics.. (25 features)

Finally, we have a list of 35 features for all the training and test images.

5.2 Classification

As mentioned before, The classification task is hard due to the low contrast of skin lesions and the high degree of visual similarity between melanoma and non-melanoma lesions. Besides, imbalanced data makes it even harder. Thus, in order to have a better balanced accuracy i used a forward variable selection method based on SVM to keep only features with balanced accuracy greater than 0.55. The variables selected by this method are :

- Assimmetry of color intensity (f3)
- Difference between mean values of the a^* component (CIEL a^* b^* color space) in the inside and outside of the lesion (f14)
- Overlapping area of the densities for a, and b channels (f17, f18)
- Color variance(f31)
- Coarseness(f33)

After that, i used random oversampling technique in order to have a balanced dataset.

Table 2: Results

	AC	MCC
SVM	0.69	0.36
Decision Trees	0.66	0.30
Bagging	**0.71**	**0.42**
Random Forests	0.67	0.33

For all the methods and in order to find the optimal hyper-parameters i used cross validation with grid search as we did in the practical session of IMA205.

As we can see from the table Bagging gave the best results, Bagging makes use of many decision trees in order to lower the variance and build a strong classifier, so it is expected that it performs better than a single decision tree. On

the other hand random forests which has some similarities with bagging is not as efficient in our case, this can be explained by the number of features selected, we have only six features and since random forest ignores some features in the choice of best split it will not perform very well, However, if we don't perform variable selection and use all the features available random forests would give better results.

References

[1] Qi Dou al. Lequan Yu, Hao Chen. Automated melanoma recognition in dermoscopy images via very deep residual networks. 2016.

[2] Thomas R. Schopfb Kevin Thonb Marc Geilhufea Kristian Hindberga Herbert Kircheschc Kajsa Møllersenb Jörn Schulza Stein Olav Skrøvsethb Fred Godtliebsena Maciel Zorteaa, . *Artificial Intelligence in Medicine*, 2014.