Differentiating Between Psoriasis and Eczema

Peter Lunding Jensen, s162869 and Emilie Knudsen Brun, s162884

Abstract

In this paper we provide a solution of how to handle large images containing skin deceases and how to differentiate between these. We further provide a tool (homepage) such that non-experts in machine learning still can benefit from this solution.

1 Introduction

Over the recent years the advances in computer technical skills in the field of deep learning has made it possible to achieve groundbreaking results in image classification7. An interesting field in which to apply the new found techniques is the medical industry. This paper will focus on trying to differentiate between pictures of Psoriasis and Eczema.

This paper is based on images depicting different genres of said deceases. The pictures used for training and testing were provided by Kenneth Thomsen from the dermatological department at the Aarhus Universityhospital (AUH). This is not a public available dataset.

Given large images as input, we are challenged by the computers ability to keep such large images in memory if they were to be given to a vanilla classification network, we therefore use a Spatial Transformer Network (STN) to zoom into the area of the provided images that are of interest to us. The output of the STN is of such a size that we are able to keep it in memory together with a classification network.

2 Spatial Transformer Networks

The images discussed were provided in a very high resolution. Hence the actual size, the amount of data so to say, were very large, and an approach to shrink the size without too high loss of information was necessary. By implementing a spatial transformer network, it is possible to provide a way in which to make the computer find the area of interest. If given enough data, the spatial transformer network will be able to adjust and improve numeral traditional state-of-the-art image classification neural networks, by providing a pre-layer resulting in an affine transformation of the image, before it is send on to the classification network (7). In our implementation, the size of the images the spatial transformer network takes as input is 630×945. The output image is of size 210×315

How it Works

The spatial transformer network implemented for this problem is inspired by [1]. The network is essentially split in three parts. A localization network, which identifies the important aspects of the image, a grid generator, which is essentially the affine transformation of the area of interest, and finally the sampler, which reconstruct a readable output from the original picture, sampled at the points determined by the grid generator (7). The final product of this process is readable by a CNN. The implementation of the code found its inspiration from 7.

The spatial transformer implemented can be shown to zoom in, in the middle of the picture, with slight variations. This is due to overfi ting. Early stopping has shown signs of improving the setup by resulting in a more divergent zoom.

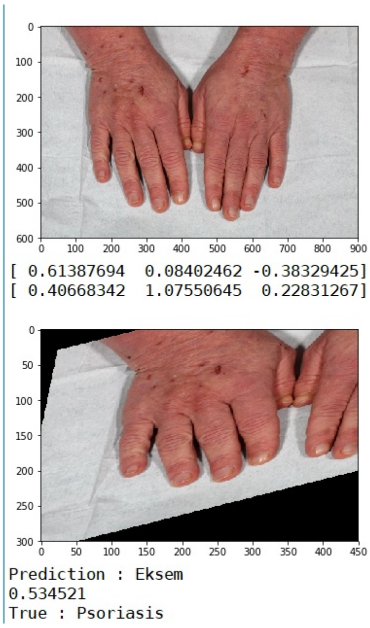

Figure 1: Example of STN while training. The matrix describes the zooming, skewing and x-y displacement compared to the original image

3 Classification Network

The VGG16 network is a well studied and well understood classification network, 7, 7, we therefore draw inspiration to our classification network from VGG16. The architecture (with modifications) is as follows:

Input layer conv(feature maps = 64, kernel = (3,3)) conv(feature maps = 64, kernel = (3,3)) maxpool(kernel=(2,2))
conv(feature maps = 64, kernel = (3,3)) conv(feature maps = 64, kernel = (3,3)) maxpool(kernel=(2,2))
conv(feature maps = 128, kernel = (3,3)) conv(feature maps = 128, kernel = (3,3)) conv(feature maps = 128, kernel = (3,3)) maxpool(kernel=(2,2))
conv(feature maps = 256, kernel = (3,3)) conv(feature maps = 256, kernel = (3,3)) conv(feature maps = 256, kernel = (3,3)) maxpool(kernel=(2,2))
dense1(2048) dropout(dense1) dense(2058) dropout(dense2) dense(Number of skin diseases)

Where we have made slight alterations compared to the VGG16 network in order to be able to keep both the STN and the classification network in memory simultaneously. Most notably is the reduction in feature maps, and that we are dealing with a binary classification problem where the original VGG16 network is classifying 1000 different classes.

The input to our classification network is the output from the spatial transformer network. In our case the output of the STN is 210×315.

4 Training

In order to keep both the STN and the classification network in memory simultaneously we have to use a small batch consisting of 15 images. We use a Adam optimizer with a learning rate of 0.00001. This learning rate is chosen in order for the STN not to diverge (e.g returning a black image). We trained the model over the course of 3 days, with interruptions (saving the model, and load it later and train from there on). The results obtained in this paper comes from approximately 10.000 batches. The graphics card used was a gtx 1060.

5 Homepage as tool for non ML experts

We have created a homepage such that non experts in machine learning (e.g. doctors) can use this model to differentiate between Psoriasis and Eczema. The homepage was created using flask version 0.12.2. To download and the tool go to www.github.com/asdkfasdfsadfasdf

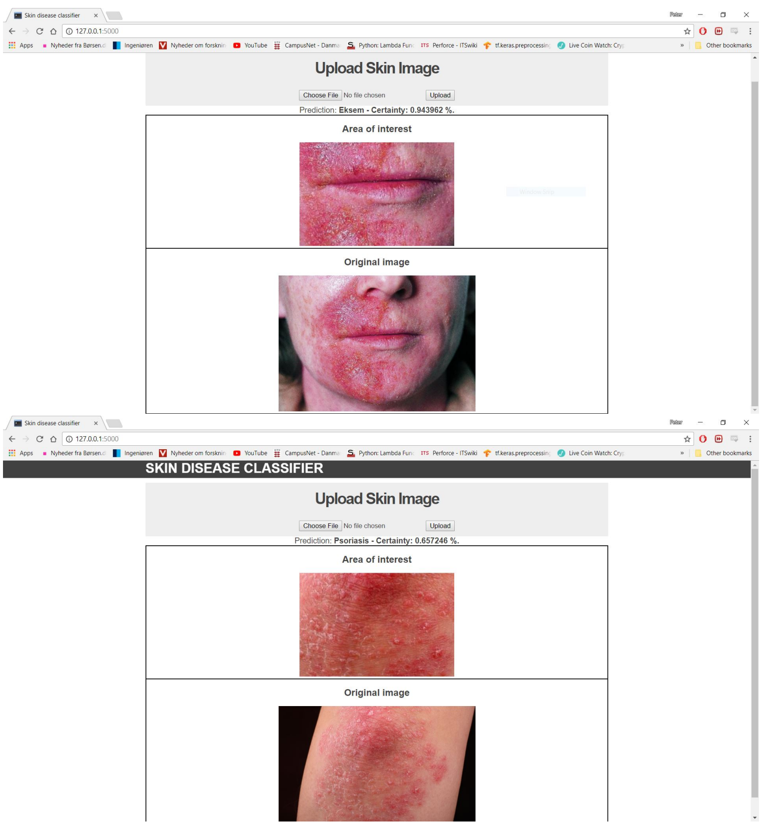

Figure 2: Screenshot from the homepage, the Area of interest is the output from the STN

6 Conclusion and results

When evaluating our model on a previously unseen data set containing 200 pictures, 100 of eczema and 100 of psoriasis, we get a performance of 67.3%. This is in itself good, but if we make a sharper cutoff, and only accept pictures for which the model is 95% certain of a value, we are able to classify approximately 90% of the pictures correctly. See fig. 3

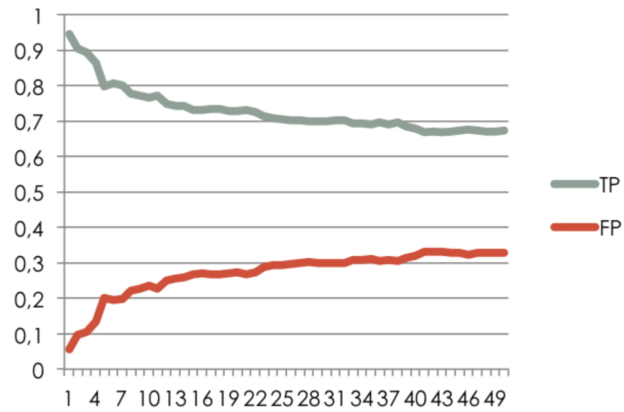

Figure 3: True positive vs False positive, the y-axis is how sure is the network is in its prediction.

We have shown that it is possible to teach the computer to differentiate between psoriasis and eczema. However, better results could be obtained if the images used for training were taken such that only the decease were in the image.

7 references

Ian Goodfellow and Yoshua Bengio and Aaro Courville
Deep Learning publisher=MIT Press, `http://www.deeplearningbook.org`, year2016

Max Jaderberg and Karen Simonyan and Andrew Zisserman and Koray Kavukcuoglu
Spatial Transformer Networks, year 2015, http://dblp.org/rec/bib/journals/corr/JaderbergSZK15

https://github.com/skaae/recurrent-spatial-transformer-code

Karen Simonyan and Andrew Zisserman
title Very Deep Convolutional Networks for Large-Scale Image Recognition,2014, http://arxiv.org/abs/1409.1556, http://dblp.org/rec/bib/journals/corr/SimonyanZ14a