Capstone Proposal

Willian Ver Valen Paiva

1 Machine Learning Engineer Nanodegree

1.1 Context

For a long time work in facial recognition has being done and one of the key points of this work is the face alignment as it poses its own challenge. And the main tool used for the job is OpenCV which is used with DLIB to recognize Facial landmarks.

The automatic recognition of landmarks is essential to be able to classify facial expressions, as it makes possible to classify Facial Action units also known as FACS [1], which in turn is used to recognize facial expression and sentiment analyses of such expressions.

As one of my main projects today is to create a DNN capable to recognize facial expression of pain, this subject comes to be perfect as it cover a personal necessity and brings a good subject to work and learn.

1.2 Existent Solutions

Today we have many different implementations to automatically recognize facial landmarks on images. But yet find an implementations on Tensorflow is not that easy. As most of the works done on the subject is heavily dependent of DLIB to recognize the landmarks.

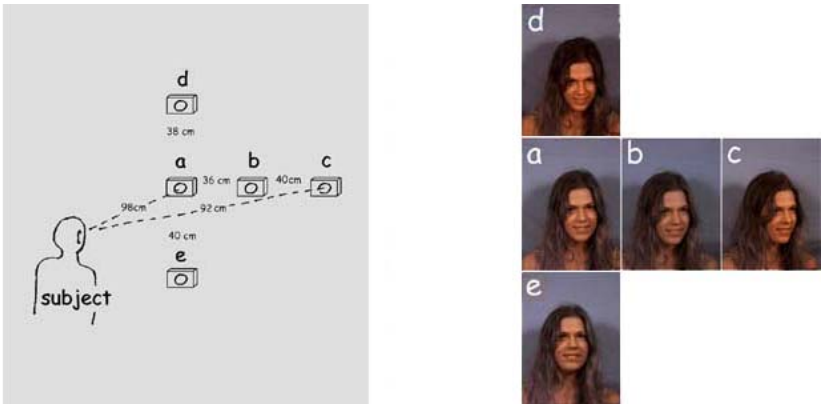
Even the project Facenet [2] uses DLIB for this task. while some projects like the 3D facial landmarks [3] from Adrian Bulat provide remarkable results it is implemented in torch.

So for that reason i propose for this project to create a Deep Neural Network to tackle such subject.

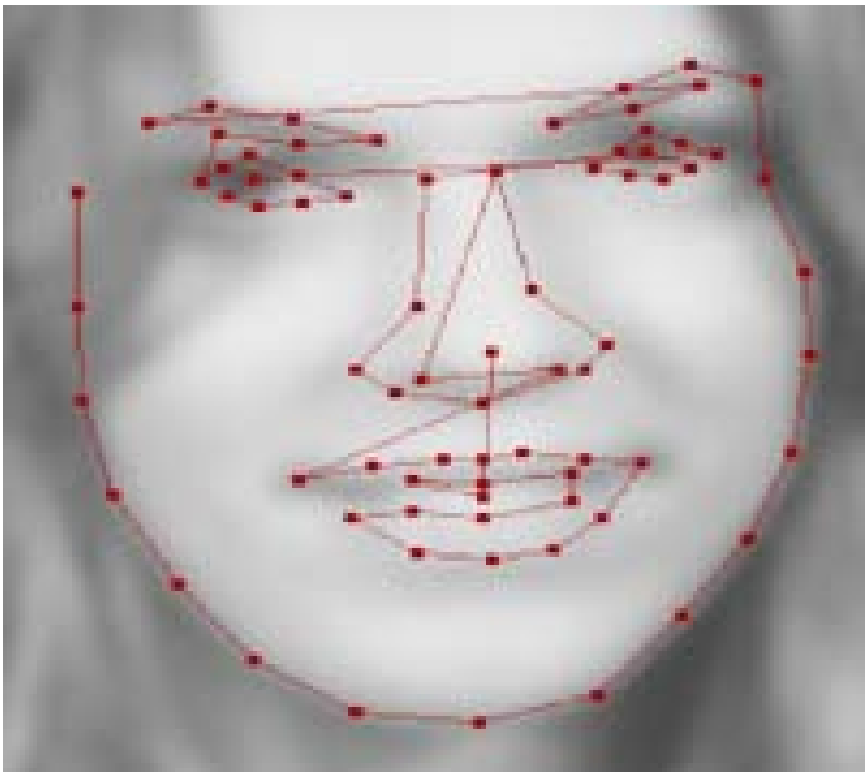
1.3 The MUCT data base

To work out this problem i propose to use the MUCT Face Database [4] this dataset consists of pictures taken from 276 subjects using 5 cameras

in different angles an light conditions like the image below: #+CAPTION There is no images on the left but they cam be reproduced by mirroring the right side



Each picture is coded with 76 facial landmark like:



The dataset is public available via github on the following link `https://github.com/StephenMilborrow/muct`

1.4 Solution and road map

To solve such a problem i will begin from the point of training my own model using Convolutional Neural Networks and compare the results by using transfer learning from other models inception v3 , resnet, etc.

The main idea here is to create a model with 152 regression outputs giving the respective X and Y of each point.

In case the transfer learning don't give good results another approach would be go up on the pre-trained model and get more fine tuning. By using the option:

```
include_top=True
```

when importing the pre-trained model in Keras, and by freezing different parts of the model is possible to archive diferent results.

1.5 Evaluation and Metrics

As the problem consists on a regression model i believe that for the evaluation of the results we could use the accuracy calculated by using one of the regression functions R-squared OR mean-squared.

References

[1] Paul Ekman and Wallace V Friesen. Facial action coding system. 1977.

[2] David Sandberg. facenet. `https://github.com/davidsandberg/facenet`, 2017.

[3] Adrian Bulat and Georgios Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In *International Conference on Computer Vision*, 2017.

[4] S. Milborrow, J. Morkel, and F. Nicolls. The MUCT Landmarked Face Database. *Pattern Recognition Association of South Africa*, 2010. `http://www.milbo.org/muct`.