Detection of Facial Landmark Attributes in Human and Non-Human Primates

DD2424 - Deep Learning in Data Science - Project Proposal

Niels Agerskov	Samy Ahmed	Graziella El-Ghorayeb
agerskov@kth.se	samyahmed9@gmail.com	graeg@kth.se

The problem that we choose to work with is the detection of five facial landmarks as well as detection of whether the person is smiling or not. We choose to work with the MTFL dataset provided by [1]. The problem will be solved by retraining the final layer of a pre-trained convolutional network, Inception V3. The fully connected layer at the top of the network will be removed and a new fully connected layer will be designed and trained to solve the chosen problem using the MTFL dataset.

The MTFL[2] dataset will be used for training, validation and testing by splitting up the data into smaller datasets. The MTFL dataset is good since it consists of a large number of images labled with both of the features that are going to be detected. In addition to the data provided by the MTFL dataset, the trained model will also be tested on non-human primate faces, provided by the PrimFace[3] dataset. This will give an idea about how the model generalises to similar but not identical facial structures.

The model will be built using the TensorFlow API, possibly with Keras as a high-level wrapper. From the API a pre-trained model of the Inception V3 network will be used. The top layer of the pre-trained network will be removed and retrained to our specific data.

The initial experiments will consist of:

- Varying the size of training, validation and test sets.

- Varying the learning rate, regularisation and complexity (of the final layer).

- Recreating landmark and smiling detection similar to [2].

- Test the network generalisation to the PrimFace dataset.

The project will be considered as a success by the accuracy of the test run using labeled MTFL data. The test run is a good metric since it is on the images similar to the MTFL dataset that we want high performance. The accuracy should achieve a mean error comparable to the results of [2] not including the attribute classification, around 10%. Regarding the smiling detection, the project will be considered a success if the failure rates are comparable to those attained in [2] for landmark detection + smiling detection, about 32%. This might however not be possible since the Inception V3 network is not created for this specific task.

References

[1] Zhang Z, Luo P, Loy CC, Tang X. Facial landmark detection by deep multi-task learning. In: European Conference on Computer Vision. Springer; 2014. p. 94–108. Available from: `http://mmlab.ie.cuhk.edu.hk/projects/TCDCN.html`.

[2] Zhang Z, Luo P, Loy CC, Tang X. Facial Landmark Detection by Deep Multi-task Learning. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer Vision – ECCV 2014. Cham: Springer International Publishing; 2014. p. 94–108.

[3] Center NJ. PrimFace: Face database of non-human primates;. Available from: `https://visiome.neuroinf.jp/primface`.