Machine Learning Engineer Nanodegree
Capstone Proposal

Richard Imms

Facial Expression Recognition

Domain Background

I have chosen to focus on Facial Expression Recognition as it is an extension of work I produced during my undergrad. Since undertaking this course I have extended the knowledge I have learnt and would like to compare the results.

The difference in my knowledge know is a better understanding of tensorflow and CNN and their limitations and benefits. Coupling this with data augmentation techniques and potentially transfer learning I aim to achieve better results than I had previously.

Facial Expression Recognition is important as it can help children with learning difficulties or disabilities and health conditions to communicate more fluidly with a robotic counterpart (Belpaeme, T. et.al).

Problem Statement

The problem that needs to be solved is the ability to recognise an emotion that is being expressed within an image. The datasets that are being used within this project are exaggerated emotions, however an extension of this work would be Facial Expression recognition in the wild.

This is a multiple category classification problem. Inputs will be a representation of an image in the format of a Numpy Array. The image will be an image of a person representing an expression.

I will be tackling this as a image classification problem using a CNN to correctly classify the expression being expressed in the faces.

Datasets and Input

The common dataset within this Domain is the Cohn-Kanade dataset, however there are other datasets such as JAFFE. Both of these datasets are exaggerated datasets in that the emotion is non complex . An example of a complex emotion would be angry and sad. This network focuses on discrete emotions such as happy, angry, sad, confusion, but never a combination.

We can combine and normalize the inputs from multiple data sources and then split them into training, validation and verification datasets. The mixture of datasets will hopefully remove any bias towards race and gender.

To prepare the dataset we will rotate all images around their eyes which will help to center the images and then we will scale them so that the inputs are the same size. Removing backgrounds from the images and leaving just the facial landmarks will be useful in improving the network so that it recognises that our interest is in the expression of these landmarks and not the background images etc.

Solution Statement
A convolutional neural network will be used to both train and test the performance of the network. By using the above data augmentation techniques the problem domain moves from one of data to one of model improvement. Using the model improvement techniques, such as k cross fold validation and batch learning I aim to find the optimal solution towards this domain. I will attempt to both build the network structure myself to help train the network and will compare it to results of transfer learning. Currently I believe transfer learning may provide better results because, even with the data augmentation techniques, the dataset is still considerably smaller than one would like for a CNN project.

Benchmark Model
The benchmark model I will be using will be the other attempts of this experiment with different models, these have been reasonably outlined in Elena Sonmez's paper on Facial Expression Recognition. (Sonmez, Elena). The performance metric used here is how well the network is at predicting the testing set for each different emotion.

Evaluation Metrics
During training and validation the labels and performance will be that supplied by the dataset. The aim is to create a general recognition and not overfit to the datasets we provide. To do this we keep an eye on the validation set and ensure that the training set doesn't rocket to close to 100% and the validation set has little to no improvement on random guessing.

As this isn't a new domain I will compare my results to the current state of the art.

Project Designs

Before beginning the design of the network I will review the current state of the art papers and see which models and data preprocessing approaches they are using. Preparing the data will probably take the most amount of time.

Once the dataset is prepared in one format I will begin constructing similar models to those outlined in previous research papers in the aim to try and reproduce their results. Once reached I will attempt different network configurations and add some more data augmentation techniques to see if there is a way to improve on the current state of the art.

Finally, I will attempt to compare the results from a ground up model to one of transfer learning.

References

Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., Racioppa, S., Kruijff-Korbayová, I., Athanasopoulos, G., Enescu, V. and Looije, R., 2013. Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, *1*(2), pp.33-53.

Sönmez, E.B. and Cangelosi, A., 2017, March. Convolutional neural networks with balanced batches for facial expressions recognition. In *Ninth International Conference on Machine Vision (ICMV 2016)* (Vol. 10341, p. 103410J). International Society for Optics and Photonics.