Winter Internship Project Report

on

Real Time Age, Gender and Emotion Detection using Deep Learning Techniques

LeadingIndia.ai

Submitted To:
Dr. Madhushi Verma
Assistant Professor,
Department of Computer Science
Miss Divya Acharya
PhD Scholar

Submitted by Team Number 1:
Reethesh Venkataraman
Amrita University
Dinesh Rochiramani
INSOFE
Manish
SRM Institute of Science & Technology
Falak Bhardwaj
MRIIRS

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
SCHOOL OF ENGINEERING AND APPLIED SCIENCES
BENNETT UNIVERSITY
GREATER NOIDA
UTTAR PRADESH

Table of Contents

CONTENT	PAGE NUMBER
Abstract	Page 3
Introduction	Page 4
Why and how the project was chosen	Page 4
Literature Survey	Page 5
Project Design Diagram	Page 6
Functionality of the Project	Page 7-8
Implementation Details	Page 8-11
Conclusion	Page 12
Limitations	Page 12
Future Enhancements	Page 12
References	Page 13
Project Links	Page 13

Abstract

The papers (given below in Literature Survey) describe the details of age, gender and emotions recognition systems. The backbone of our system consists of several Deep Convolutional Neural Network (CNN). To power these networks, we utilized a large labelled dataset through a semi-supervised pipeline to reduce the annotations efforts and time. Age, gender and facial expression classification has become relevant to an increasing amount of applications, particularly since the rise of social media. Nevertheless, performance of the existing methods on the real-world images has scope of improvement, especially as compared to the tremendous leaps in performance recently reported for the related task of Face Recognition. In these papers, it is claimed that by learning representations using Deep Convolution Neural Network, a significant increase in performance can be obtained on these tasks. We proposed a simple convolutional net architecture that can be used easily.

Introduction

Emotion is a strong feeling deriving from one's circumstances, mood, or relationship with others. It tells us about one's social behaviour and presence of mind. Social etiquettes play a major role in one's personality.

A growing number of applications rely on the ability to extract information about people from real time input through the camera. Examples are the person identification for surveillance or access control the estimation of gender and age that have been addressed in isolation in the past, there often exists a variety of methods for each.

This project was done in the context of building a program with different modules which takes real time inputs using a webcam and detects information such as age, gender and emotion. This chosen topic is one of great social interest. Furthermore, our goal was to create a universal solution to this problem.

Emotions that we have worked upon to carry out the experiment are 7 in total:

1. Anger
2. Disgust
3. Happy
4. Sad
5. Surprise
6. Fear
7. Neutral

Why and how the project was chosen

For our pre-project research, we were asked to go through different papers and journals being published in recent years related to Deep Learning, for getting a basic idea of which topic were going to pursue as our project. We decided to follow our mentor's advice to take up a project with a contemporary field.

During our study, we kept in mind the need to tackle real life problems. One such problem was the reduction of mortality due to mental instability and suicides. Such people, in need of help, never happen to reveal their feelings. Instead, their expressions are our only way of deduction. Nowadays, surveillance is present in almost every corner in the form of a CCTV camera, or even a simple computer webcam. We decided to put those to good use.

Our mentor had talked to us about the Big Five personality traits, shortly abbreviated as OCEAN. When taken as a tree, OCEAN formed the parents, which had facets and subsequent sub-facets. In the end, they could be expanded to 27 basic emotions. We took a bottom up approach to the problem and decided to train our model to identify 7 basic emotions out of the 27.

This project was one step towards Personality detection, which can be furthermore worked upon in the future for effective personality recognition and be used for the greater good. It can further be used to identify one's past doings (if recorded) and present, if the Neural Network is trained on well labelled dataset. So, it can be useful to identify sociopaths with bad intentions as well.

Literature Survey

We began our background search with research papers and blog posts online, related to our topic. The research paper details:

This paper tackles the estimation of apparent age in still face images with deep learning. Their CNNs use the VGG-16 architecture and are pretrained on ImageNet for image classification. They crawled 0.5 million images of celebrities from IMDB and Wikipedia that were made public. They pose the age regression problem as a deep classification problem followed by a softmax expected value refinement and show improvements over direct regression training of CNNs. Their proposed method, Deep EXpectation (DEX) of apparent age, first detects the face in the test image and then extracts the CNN predictions from an ensemble of 20 networks on the cropped face **[1]**

This paper presents automated, real-time models built with machine learning algorithms which use videotapes of subjects' faces in conjunction with physiological measurements to predict rated emotion. They built algorithms based on extracted points from the subjects' faces as well as their physiological responses. Results demonstrated good fits for the models overall, with better performance for emotion categories than for emotion intensity, for amusement ratings than sadness ratings, for a full model using both physiological measures and facial tracking than for either cue alone, and for person-specific models than for gender-specific or general models **[2]**

Project Design Diagram

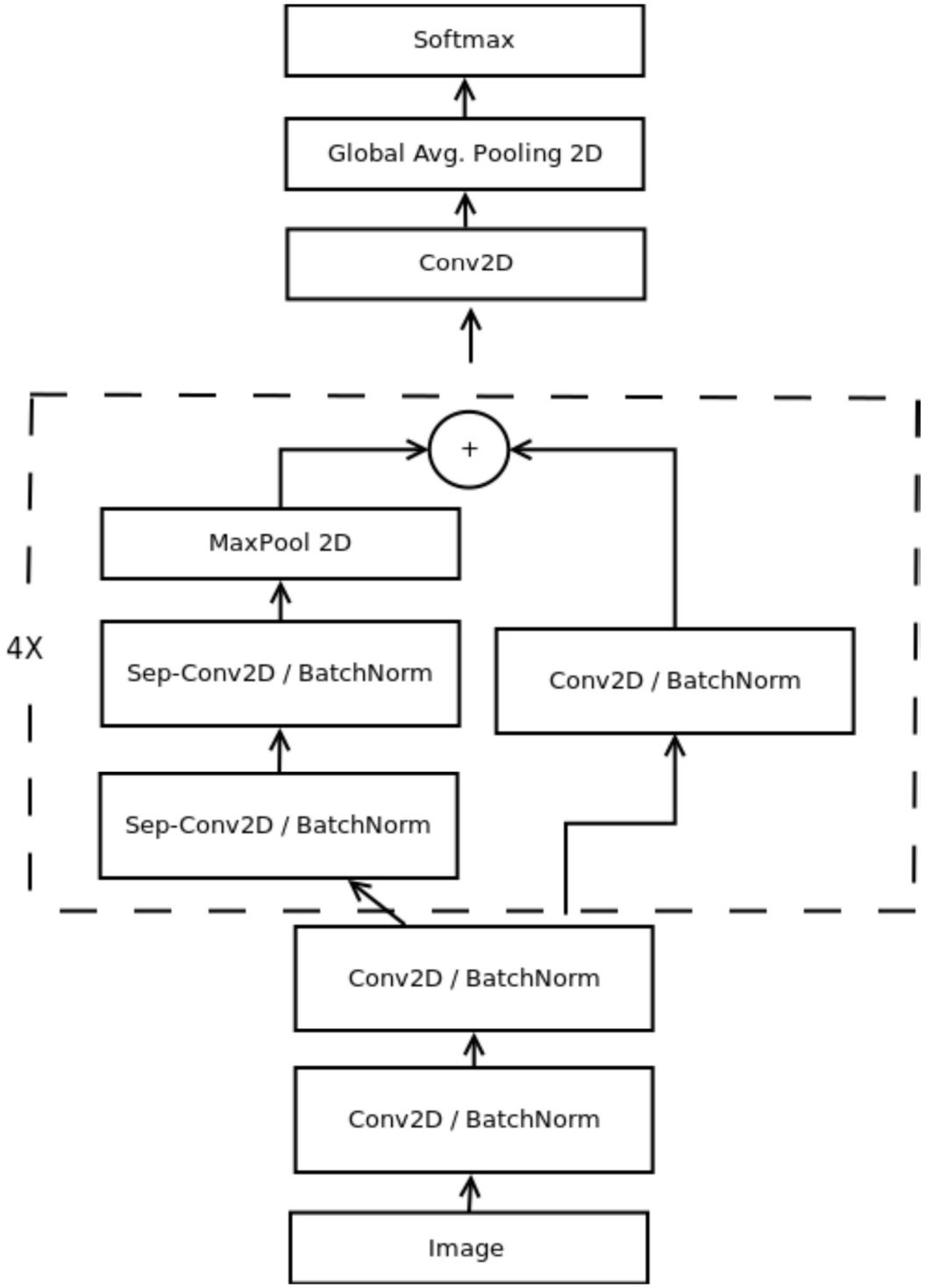

Figure 1: Emotion Detection Flowchart

Functionality of the Project

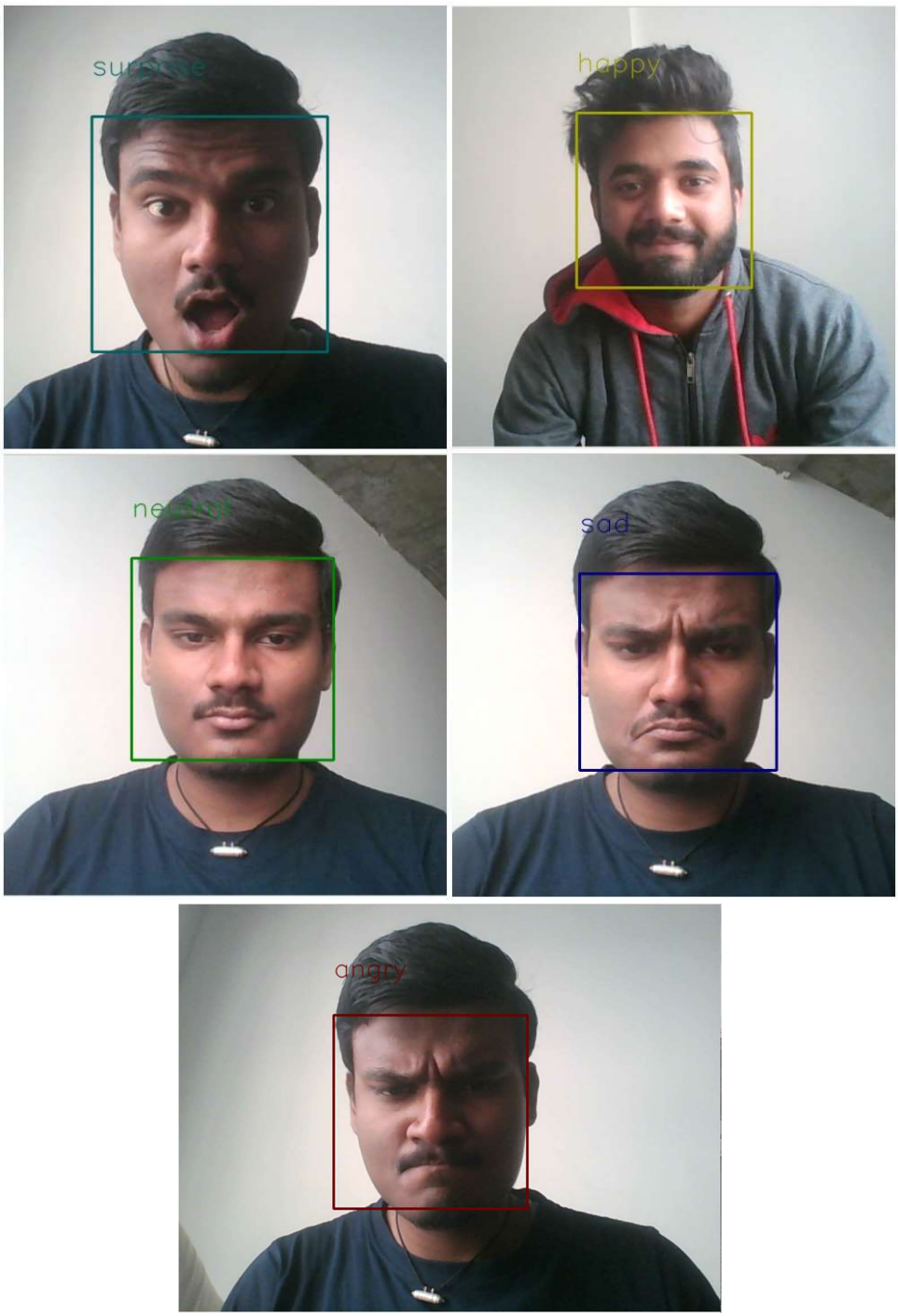

Figure 2: EMOTIONS COLOR DEMO

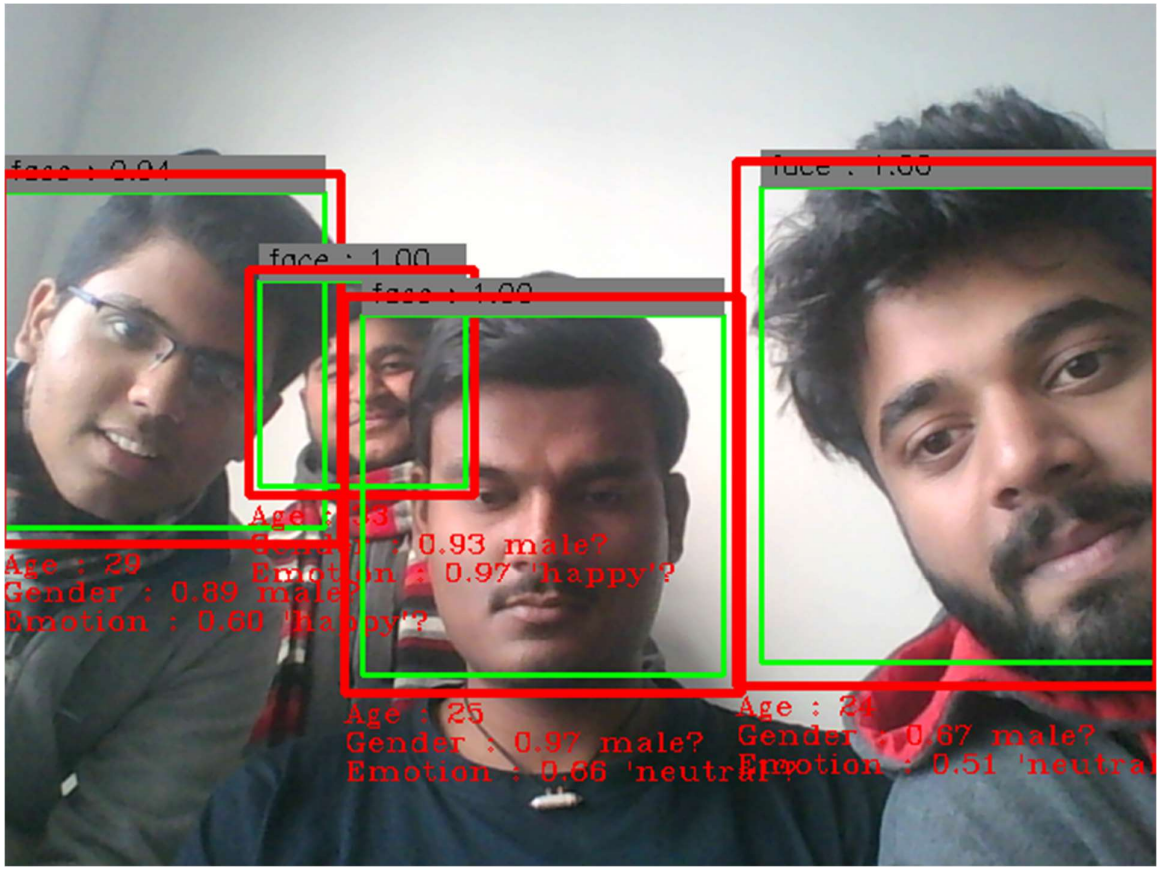

Figure 3: COMPLETE PROJECT DEMO

Implementation Details

Prerequisites are:

- Keras2 (with TensorFlow backend)
- OpenCV
- Python 3.5 (TensorFlow not supported in higher versions)
- Darknet (YOLOv2 for face detection training, YOLOv3 can be used).
- Pillow (for rendering test results)
- NumPy
- TensorFlow
- h5py (for Keras model serialization)

Datasets used are:

- FDDB (Face Detection Database and Benchmark) – for face detection
- IMDB-WIKI – for age and gender detection
- UTK Face – for age and gender detection
- FER 2013 – for emotion detection

All datasets are first annotated. After annotation, 0 in the label stands for male and 1 stands for female. The current age or the person is calculated by subtracting his birthdate from the date the picture was taken.

Face detection is trained using Darknet, more specifically, the "YOLOv2-tiny" model. You only look once (YOLO) is a state-of-the-art, real-time object detection system. On a powerful GPU, it can process images at 30 FPS. We can easily tradeoff between speed and accuracy by simply by changing the size of the model, with no retraining. YOLO applies a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities. One main advantage is that it looks at the whole image at test time, so its predictions are informed by global context in the image. This makes it 1000x faster than R-CNN. YOLOv2-tiny is a small model for constrained environments. After training, it is converted into a Keras model with the YAD2K conversion script by GitHub user allanzelener.

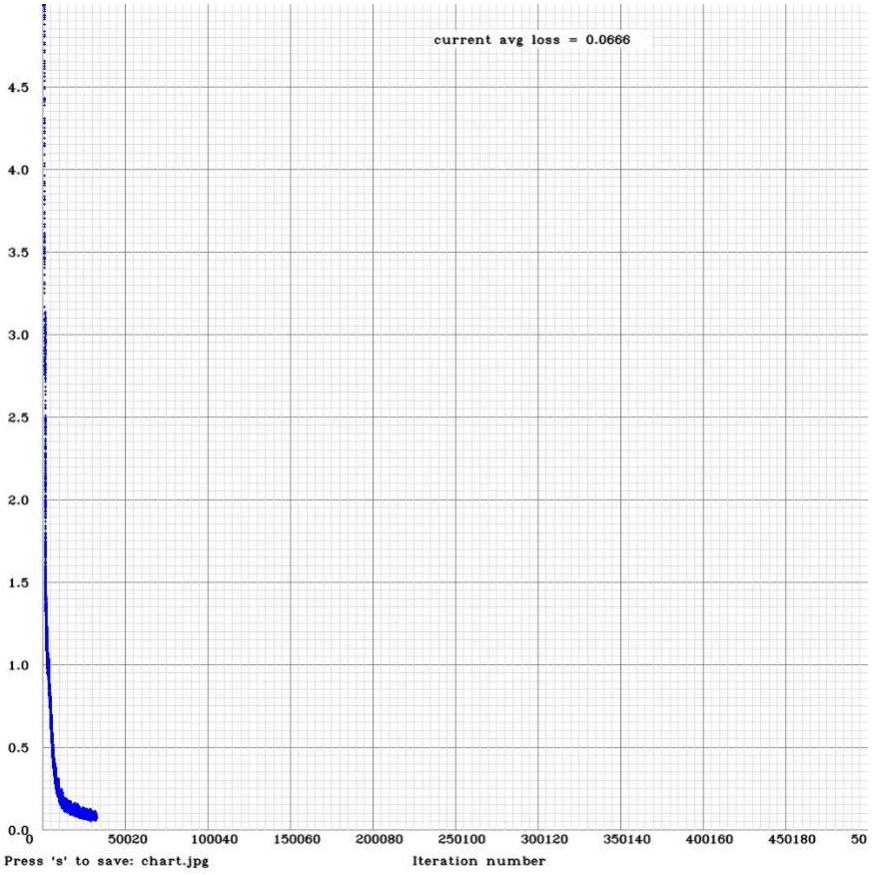

Figure 4: FACE DETECTION TRAINING RESULT

Age and Gender classification is trained using the Squeezenet network model. It is an implementation of the Keras Functional Framework 2.0. It has AlexNet accuracy with a much smaller footprint (510x smaller than AlexNet) and greater speed (approximately 4x). Pretrained models are converted from original Caffe network. Several fire modules are implemented with expand1x1 and expand3x3 layers which are concatenated together in the channel dimension.

Figure 5: AGE DETECTION TRAINING RESULT

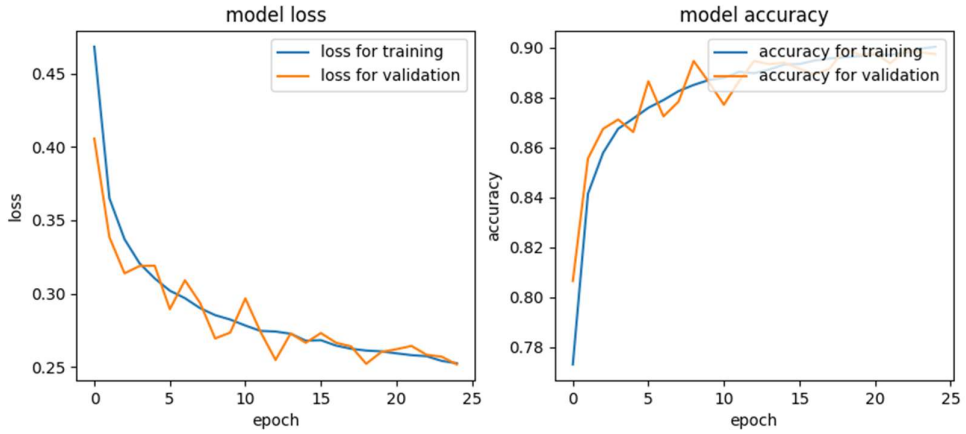

Figure 6: GENDER DETECTION TRAINING RESULT

Our emotion detection model is inspired by the Xception architecture. This architecture combines the use of residual modules and depth-wise separable convolutions. Residual modules modify the desired mapping between two subsequent layers, so that the learned features become the difference of the original feature map and the desired features. The architecture is trained with the Adam optimizer. Depth-wise separable convolutions are composed of two different layers: depth-wise convolutions and pointwise convolutions.

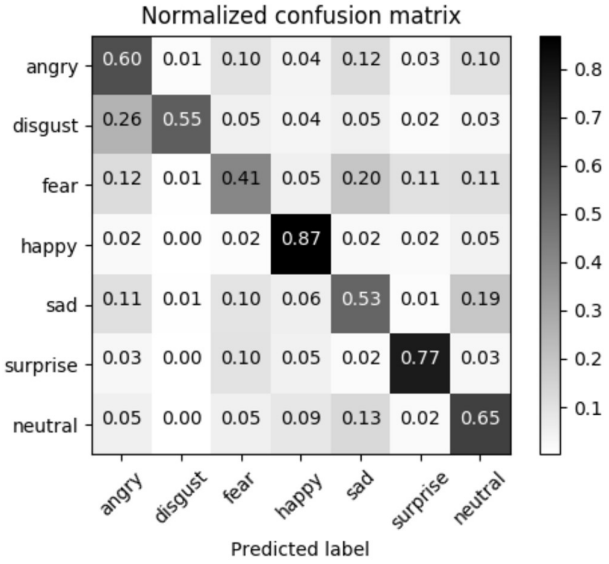

Figure 7: NORMALISED CONFUSION MATRIX OF MINI-XCEPTION NETWORK

Conclusion

We have proposed and tested a general building designs for creating real-time CNNs. Our proposed architectures have been systematically built in order to reduce the number of parameters as much as possible. We have shown that our proposed models can be stacked for multiclass classification while maintaining real-time inferences.

In conclusion, we've successfully constructed a working CNN model to recognize the Facial Expressions, Age and Gender of Human Beings.

The accuracy achieved for the different models on training are as follows:

- Age detection – 7% (prediction from age 0 to 100, trained with a significantly large dataset)
- Gender detection – 93%
- Face detection – 96%
- Emotion recognition – 70%

Limitations

Some of the major limitations that we faced during the development of the project are:

- The training of the networks requires the latest, large and different types of datasets.
- Training requires high computation power due to the large size of the dataset and high number of parameters for the CNN models.
- The post deployment results will be near real time only, mainly due to the low FPS from the webcams.

Future Enhancements

Machine learning models are biased in accordance to their training data. In our specific application we have empirically found that our trained CNNs for age and gender classification are biased towards western facial features and facial accessories. We hypothesize that this misclassification occurs since our training dataset consist of mostly western: actors, writers and cinematographers. Furthermore, as discussed previously, the use of glasses might affect the emotion classification by interfering with the features learned. However, the use of glasses can also interfere with the gender classification. This might be a result from the training data having most of the images of persons wearing glasses assigned with the label "man". We believe that uncovering such behaviors is of extreme importance when creating robust classifiers, and that the use of the visualization techniques such as guided back-propagation will become invaluable when uncovering model biases.

References

[1] Bailenson, J. N., Pontikakis, E. D., Mauss, I. B., Gross, J. J., Jabon, M. E., Hutcherson, C. A. C., … John, O. (2008). **Real-time classification of evoked emotions using facial feature tracking and physiological responses**. International Journal of Human-Computer Studies, 66(5), 303–317.

[2] Rothe, Rasmus, Radu Timofte, and Luc Van Gool. "**Dex: Deep expectation of apparent age from a single image**." In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 10-15. 2015.

[3] Celiktutan, O., Sariyanidi, E., & Gunes, H. (2015). **Let me tell you about your personality!: Real-time personality prediction from nonverbal behavioural cues**. 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG).

[4] Kapoor, A., Burleson, W., & Picard, R. W. (2007). **Automatic prediction of frustration**. International Journal of Human-Computer Studies, 65(8), 724–736.

[5] Zagoruyko, S. and Komodakis, N., 2016. **Wide residual networks**. *arXiv preprint arXiv:1605.07146*.

[6] Dehghan, Afshin & G. Ortiz, Enrique & Shu, Guang & Zain Masood, Syed. (2017). **DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network**.

Project Links

https://youtu.be/FVhiHnNW444

https://github.com/vreethesh/AGE---Age-Gender-Emotion-detection