Indian Institute of Technology Indore

Discipline of Computer Science and Engineering

Minor Project in the course "Computational Intelligence"

Spring 2022-2023

Title: Real-time Face Detection Model with Age, Gender and Facial Emotion Prediction

Final Report

Problem Statement:

When dealing with a situation in the real world, it is crucial that we remain vigilant of everybody around us. Face detection becomes crucial as we seek information about the people present at a given location and time. The project's primary goal is to identify any human faces in front of the camera (Real-time detection of faces). Additionally, we will determine person's gender and age after detecting their face. We also determine the facial expressions of the people present in the frame.

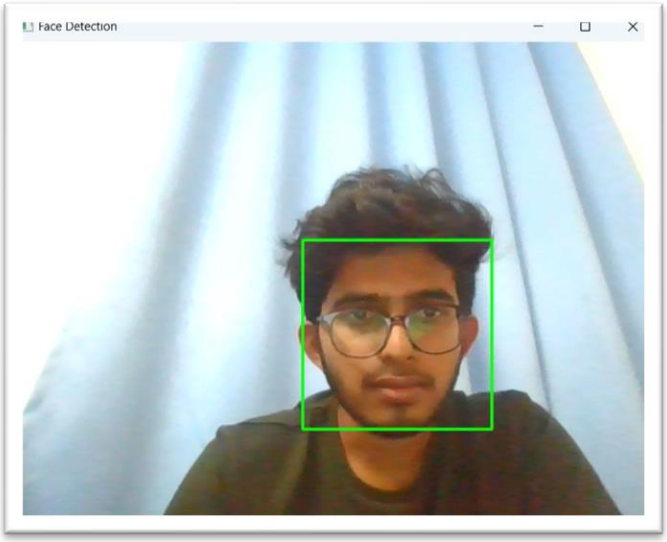
Algorithms and Outputs (Results):

Basic task is to detect the faces:

1) Algorithm: Haar Cascade Classifiers - Pre trained

Using the Haar Cascade Classifier to detect the faces present in front of the camera and surrounding the detected faces with a rectangular box.

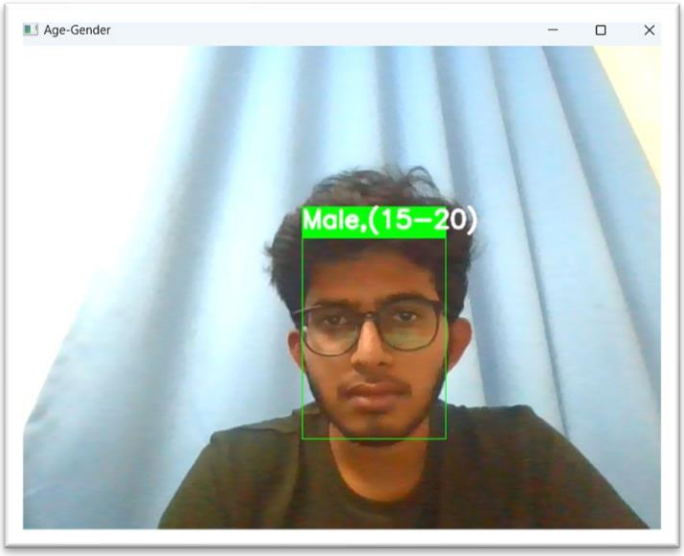
Output-



2) Algorithm: Pre-Trained Caffe Models – CNNs, FaceNet, GenderNet, AgeNet

Using pre-trained Caffe Models which are designed on CNN architecture to predict age and gender of the detected faces (detected using FaceNet).

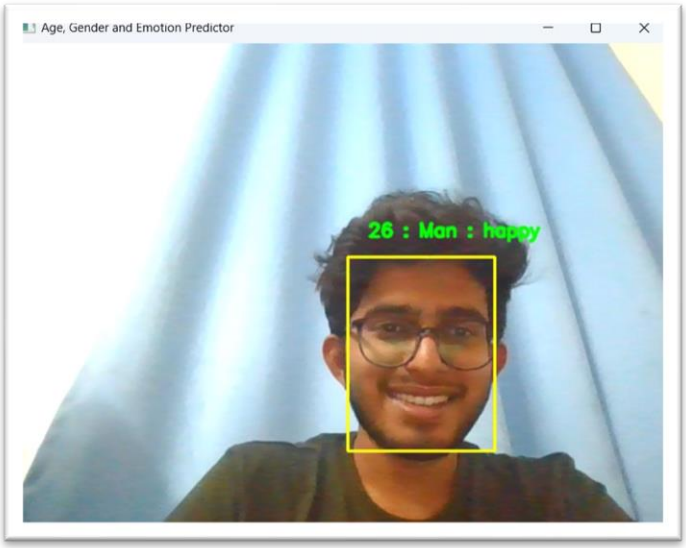
Output-



3) Algorithm: DeepFace - Age, Gender, Emotion Prediction, FaceNet - Face Detection

Using DeepFace to predict the age, gender and emotion of the individual whose face has been detected using FaceNet.

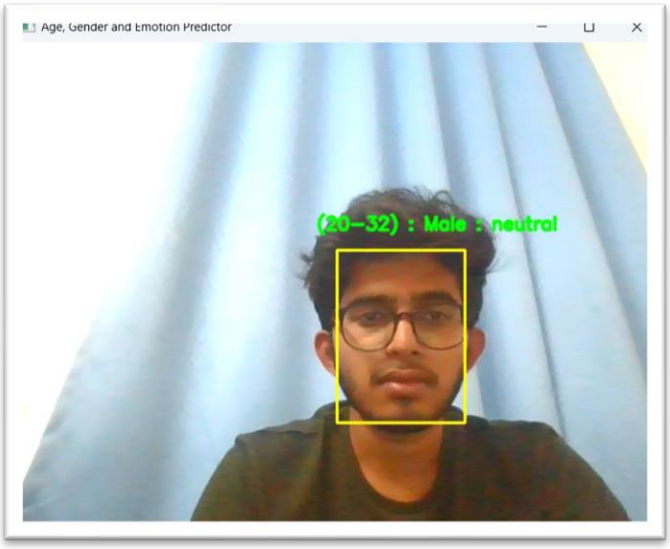
Output-



4) Algorithm: DeepFace - Emotion Prediction, FaceNet, AgeNet, GenderNet

Using DeepFace for emotion prediction and AgeNet, GenderNet for predicting the age and gender of the determined faces (FaceNet).

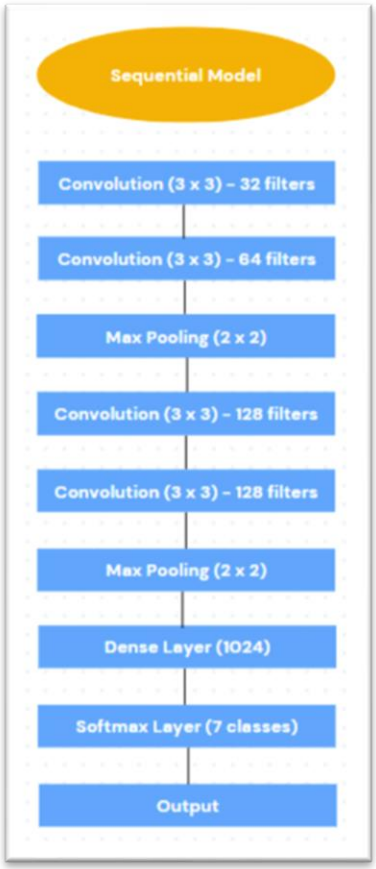
Output-



5) Algorithm: DeepFace - Age, Gender Prediction, Trained CNN - Emotion Detection

Training our own CNN Model for predicting the emotions of the individuals. DeepFace is used for age and gender prediction on the faces determined using Haar Cascade Classifiers.

CNN Model- Contains 4 Convolutional layers with kernel size (3*3) with varying filters and 2 MaxPooling layers with kernel size (2*2). Finally, a dense layer and a softmax layer – for predicting the output.



Training- The model is trained for 80 epochs with a batch size of 64. Image size – 48*48. Images are Gray scaled.

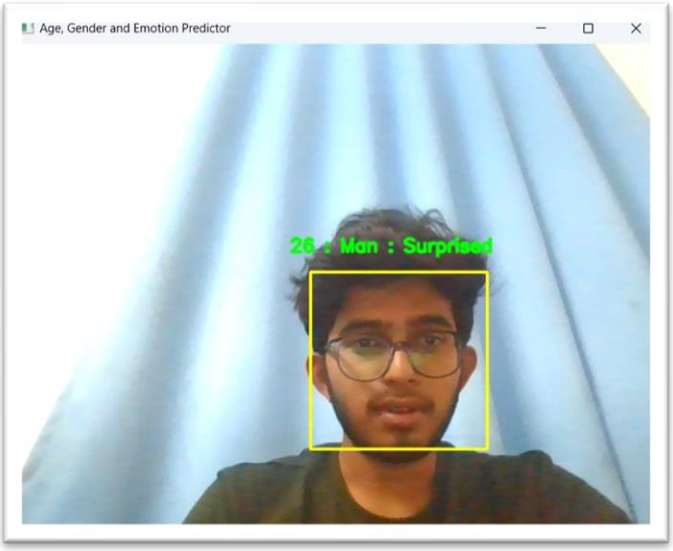
Metrics – Accuracy, Categorical Cross Entropy Loss.

Dataset - https://www.kaggle.com/datasets/msambare/fer2013

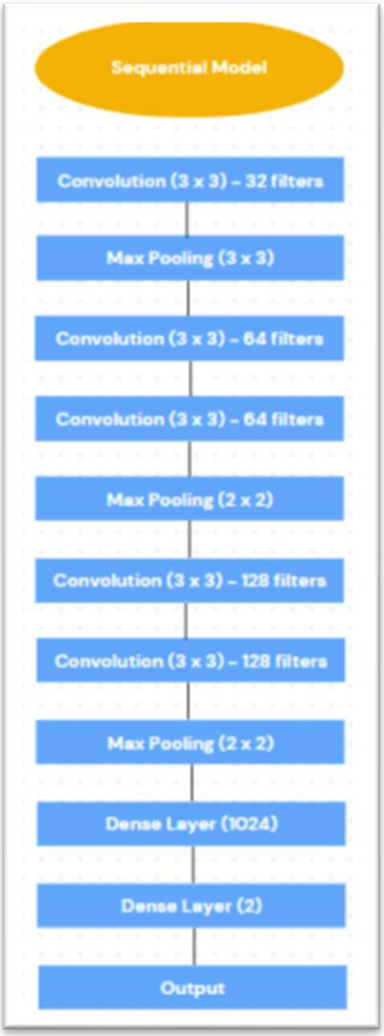
Output-

- Training Accuracy – 93.78%
- Training Loss – 0.1776
- Validation Accuracy – 62.46%
- Validation Loss – 1.4766

```
Epoch 76/80
448/448 [==============================] - 121s 270ms/step - loss: 0.1925 - accuracy: 0.9331 - val_loss: 1.4733 - val_accuracy:
0.6251
Epoch 77/80
448/448 [==============================] - 124s 277ms/step - loss: 0.1913 - accuracy: 0.9337 - val_loss: 1.4563 - val_accuracy:
0.6274
Epoch 78/80
448/448 [==============================] - 126s 281ms/step - loss: 0.1844 - accuracy: 0.9349 - val_loss: 1.4897 - val_accuracy:
0.6325
Epoch 79/80
448/448 [==============================] - 125s 279ms/step - loss: 0.1853 - accuracy: 0.9354 - val_loss: 1.4821 - val_accuracy:
0.6318
Epoch 80/80
448/448 [==============================] - 124s 276ms/step - loss: 0.1776 - accuracy: 0.9378 - val_loss: 1.4766 - val_accuracy:
0.6246
```



6) Algorithm: DeepFace - Age Prediction, Trained CNN - Emotion Detection, Gender Prediction - Model 1



Training our own CNN Model for predicting the gender of the individuals. DeepFace is used for age prediction on the faces determined using Haar Cascade Classifiers. Previously trained CNN Model is used for emotion prediction.

CNN Model- Contains 5 Convolutional layers with kernel size (3*3) with varying filters and 3 MaxPooling layers with kernel size (2*2). Finally, two dense layers - for predicting the output (Male/Female, how much accurate it is).

Training- The model is trained for 100 epochs with a batch size of 64. Image size – 96*96

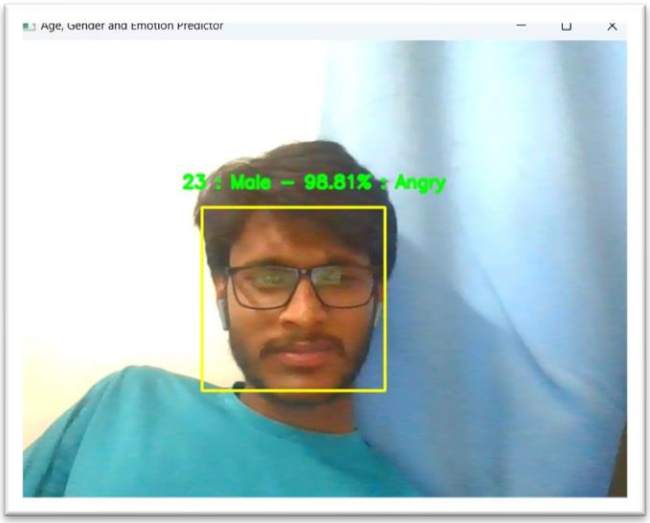
Metrics – Accuracy, Categorical Cross Entropy Loss.

Dataset - https://github.com/balajisrinivas/Gender-Detection/tree/master/gender_dataset_face

Output-

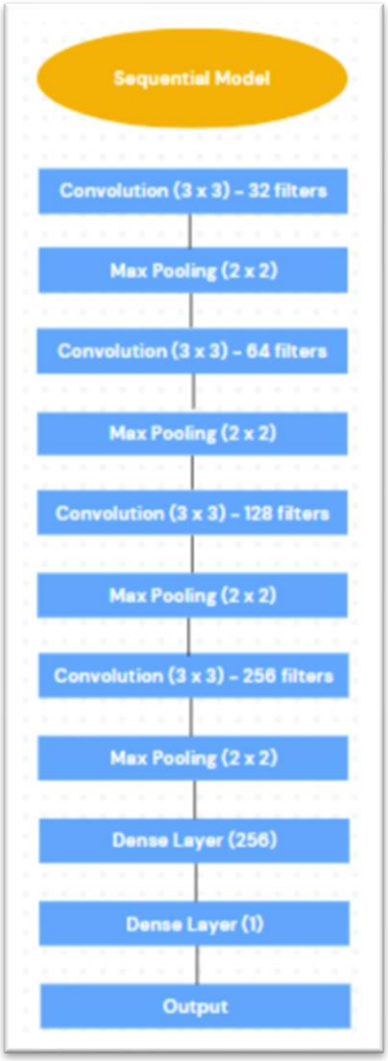
- Training Accuracy – 98.20%
- Training Loss – 0.0521
- Validation Accuracy – 93.72%
- Validation Loss – 0.3006

```
Epoch 96/100
28/28 [==============================] - 21s 730ms/step - loss: 0.0404 - accuracy: 0.9854 - val_loss: 0.1788 - val_accuracy:
0.9784
Epoch 97/100
28/28 [==============================] - 21s 746ms/step - loss: 0.0334 - accuracy: 0.9854 - val_loss: 0.1617 - val_accuracy:
0.9545
Epoch 98/100
28/28 [==============================] - 21s 728ms/step - loss: 0.0317 - accuracy: 0.9888 - val_loss: 0.0987 - val_accuracy:
0.9697
Epoch 99/100
28/28 [==============================] - 21s 731ms/step - loss: 0.0454 - accuracy: 0.9854 - val_loss: 0.1359 - val_accuracy:
0.9524
Epoch 100/100
28/28 [==============================] - 20s 696ms/step - loss: 0.0521 - accuracy: 0.9820 - val_loss: 0.3006 - val_accuracy:
0.9372
```



7) Algorithm: Trained CNN - Emotion Detection, Gender Prediction - Model 2

Training another CNN Model for predicting the gender of the individuals. Faces are determined using Haar Cascade Classifiers. Previously trained CNN Model is used for emotion prediction.



CNN Model- Contains 4 Convolutional layers with kernel size (3*3) with varying filters and 4 MaxPooling layers with kernel size (2*2). Finally, two dense layers - for predicting the output (Male/Female).

Training- The model is trained for 50 epochs with a batch size of 32. Image size – 128*128

Metrics – Accuracy, Binary Cross Entropy Loss and Mean Squared Error.

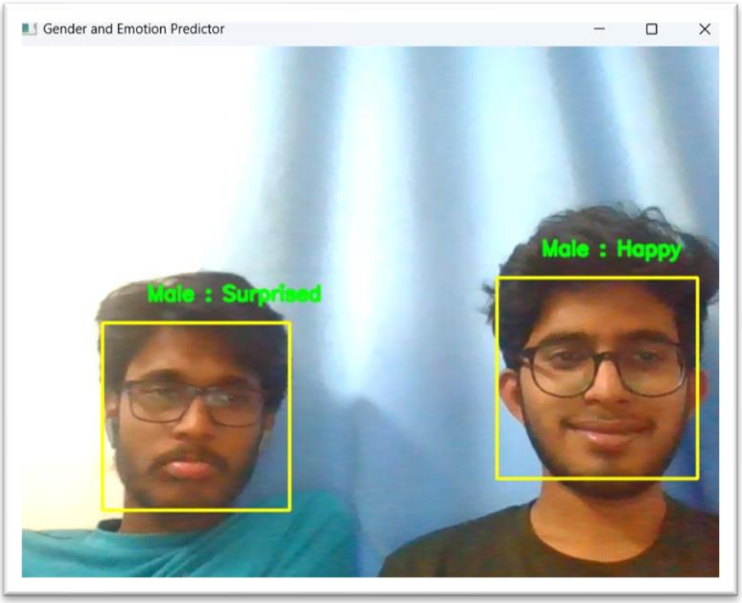
Dataset - https://www.kaggle.com/datasets/jangedoo/utkface-new

Output-

- Training Accuracy – 99.15%
- Training Loss – 0.0211

- Validation Accuracy – 82.90%
- Validation Loss – 1.5957

```
Epoch 47/50
593/593 [==============================] - 170s 287ms/step - loss: 0.0324 - accuracy: 0.9880 - val_loss: 1.4564 - val_accurac
y: 0.8330
Epoch 48/50
593/593 [==============================] - 171s 289ms/step - loss: 0.0236 - accuracy: 0.9908 - val_loss: 1.3725 - val_accurac
y: 0.8305
Epoch 49/50
593/593 [==============================] - 170s 287ms/step - loss: 0.0213 - accuracy: 0.9919 - val_loss: 1.7149 - val_accurac
y: 0.8372
Epoch 50/50
593/593 [==============================] - 170s 287ms/step - loss: 0.0211 - accuracy: 0.9915 - val_loss: 1.5957 - val_accurac
y: 0.8290
```



Analysis:

- Comparing the two models used for detecting the faces – Haar Cascade Classifier and FaceNet. Both are working efficiently. However, Haar Cascade Classifier works better in case of real-time face detection and FaceNet works well for accurate face recognition such as biometric authentication. In our case, Haar Cascade works well.
- For age and gender predictions, among DeepFace and AgeNet - GenderNet, DeepFace works more efficiently and produces accurate results.
- When compared with our trained CNN Model for emotion detection, both of them work almost at the same level with similar kinds of predictions.

- The first trained gender CNN model gives us the accuracy on how much it thinks that its prediction is true (Ex: Male – 99.02%).
- Among the two trained models for gender prediction, the first one is more accurate and gives the right predictions in most of the cases. The second one varies greatly as the face keeps moving but once it becomes kind of static, it predicts accurately.
- Our models are mostly on par with the pre-trained models.

Limitations/Challenges faced:

- The dataset that we are using for emotion prediction has images of small size (48*48). Training these images to extract the features was difficult and not many features were extracted. Also, the images are gray scaled.
- Our laptop's processing speed is pretty much low. So, the models weren't trained for huge number of epochs. So, the accuracies aren't that close to 100%. Also, the training time is too high.

Further Improvements:

- Implement other detection and predictions on the detected human faces such as race etc. Facial identification can also be performed.
- Collect and prepare a proper dataset for emotion prediction and then train the model on that dataset for better validation accuracy rather than the standard FER-2013.

References: [1] Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, *57*(2), 137-154.

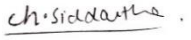
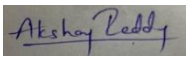
[2] Zhang, C., & Zhang, Z. (2010). A survey of recent advances in face detection.

[3] Rafique, I., Hamid, A., Naseer, S., Asad, M., Awais, M., & Yasir, T. (2019, November). Age and gender prediction using deep convolutional neural networks. In *2019 International conference on innovative computing (ICIC)* (pp. 1-6). IEEE.

[4] Arriaga, O., Valdenegro-Toro, M., & Plöger, P. (2017). Real-time convolutional neural networks for emotion and gender classification. *arXiv preprint arXiv:1710.07557*.

[5] https://github.com/balajisrinivas/Gender-Detection/tree/master/gender_dataset_face

Team Members:

Siddartha Chennareddy	200002070	
M. Sai Akshay Reddy	200001049	

Under the Supervision of

Dr. Aruna Tiwari

Professor, CSE