A Project Report on

Human age, gender and emotion recognition

Submitted in partial fulfillment of the requirements
for the award of the degree of

Bachelor of Engineering

in

Computer Engineering

by

Atharva Vaidya

16102043

Under the Guidance of

Prof. Archana Kotangale

Department of Computer Engineering
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI

Academic Year 2019-2020

Approval Sheet

This Project Report entitled ***"Human age, gender and emotion recognition"*** Submitted by ***"Atharva Vaidya"(16102043)***is approved for the partial fulfillment of the requirenment for the award of the degree of ***Bachelor of Engineering*** in ***Computer Engineering*** from ***University of Mumbai***.

Prof. Archana Kotangale
Guide

Prof. Sachin Malve
Head Deartment of Computer Engineering

Place:A.P.Shah Institute of Technology, Thane
Date:

CERTIFICATE

This is to certify that the project entitled ***"Human age, gender and emotion recognition"*** submitted by ***"Atharva Vaidya" (16102043)*** for the partial fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Computer Engineering***,to the University of Mumbai,is a bonafide work carried out during academic year 2019-2020.

Prof. Archana Kotangale
Guide

Prof. Sachin Malve
Head Deartment of Computer Engineering

Dr. Uttam D.Kolekar
Principal

External Examiner(s)
1.

Place:A.P.Shah Institute of Technology, Thane
Date:

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

(Signature)

_____

(Atharva Vaidya 16102043)

Date:

Abstract

Emotions play a very significant role in conveying what we think. Research has shown that almost 90 percent of our communication can be non verbal. The overall body language coupled with emotions could help to effectively recognize the thought process. Hence recognizing emotions becomes a vital task. There a seven types of emotions namely, happy, sad, anger, smile, neutral, disgust and fear. To effectively identify these would help to decipher the thinking. Hence the importance and need of recognizing emotions is growing rapidly. Also, age and gender play an equally important part. Recognising and effectively analysing these human characteristics is becoming important with time. This study not only finds a place in the field of security and surveillance but also form the basis of in the field where prediction of future actions is necessary.

Contents

List of Figures

List of Abbreviations

HCI:	Human Computer Interface
CNN:	Convoltional Neural Network

Chapter 1

Introduction

Emotion plays an important role in human life. Interpersonal human communication includes not only language that is spoken, but also non-verbal cues as hand and body gestures, tone of the voice, which are used to express feeling and give feedback and most importantly through facial expression. Human beings express emotions in day to day interactions. Understanding and knowing how to react to people's expression greatly enriches the interaction. The field of psychology has played an important role in understanding human emotion and developing concepts that may aid these HCI technologies.
This system is divided into two parts. The first part is about detection of face from the input frame, while the second part is about the rest of the activities comprising of emotion, age and gender recognition. Emotion recognition systems find applications in several interesting areas. With the recent advances in robotics, especially humanoid robots, the urgency in the requirement of a robust expression recognition system is evident. Emotion recognition plays a significant role in recognizing one's affect and in turn helps in building meaningful and responsive Human Computer Interface (HCI). Apart from the two main applications, namely robotics and affect sensitive HCI, emotion recognition systems find uses in a host of other domains like Telecommunications, Video Games, Animations, Psychiatry, Automobile Safety, Educational Software, etc.
Age and gender play fundamental roles in social interactions. Automatic age and gender classification has become relevant to an increasing amount of applications, particularly since the rise of social platforms and social media. This study finds an importance in the field of security and surveillance. Data that is fetched everyday is vital because a normal behavioural change is sometimes neglected and hard to identify whereas if its backed by routinely stored and analysed data the change may be quickly understood and an anomalous behaviour leading to an unwanted suspicious activity could be averted.

1.1 Problem Definition

To build an emotion, age and gender recognition system in which when system detects user's face it will be able to effectively classify user's characteristics.

Figure 1.1: 7 different emotions

1.2 Objectives

- To be able to detect the user's face and recognize human emotions, age, gender effectively.
- To be able to distinguish between the different characteristics.
- To focus on optimizing the resources.

Chapter 2

Literature Review

Images are the most direct and most natural channels that people acquire information. If achievements in these two fields are applied on robots that can great improve the intelligence of the machine. In practice, in image recognition we will encounter the feature selection problem. Common image features are composed of color feature, texture feature,shape feature,spatial relations characteristics. Sometimes in order to acquire a better final result, these characteristics also be integrated appropriately. CNN is a specially designed multi layer perceptron to identify two dimension shapes. Therefore dimensional information retained in waveform points is effectively utilized by CNN. CNN model due to its characteristics of adaptive feature extraction.

Gender and Age Classification of Human Faces for Automatic Detection of Anomalous Human Behaviour- Xiaofeng Wang, Azliza Mohd Ali, Plamen Angelov.

Human behaviour is often uncertain, and sometimes it is affected by emotion or environment. Automatic detection can help to recognise human behaviour which later can assist in investigating suspicious events. Central to our proposed approach is the recently introduced transfer learning. It was used on the basis of deep learning and successfully applied to image classification area.Automatic detection plays an important role in detecting anomalous behaviour in data and can benefit in preventing crime. The system described in may create an alert or signal if there is an anomalous behaviour. Anomalous behaviour can take place in public places such as airports, subway stations or shopping malls. In public places, there are many kinds of human behaviour. Identifying anomalous behaviour is difficult or can only be detected by means of comparison with the patterns of normal behaviour. Research on surveillance systems in public places aiming to detect abnormal human behaviour at train platforms, detecting anomalous behaviour in the crowded scene and intelligent surveillance system for public places is growing exponentially. It is not easy to access every identity in the world. Gender and age classification is one of the ways to reduce the hard work to find the identity of a human.

An Efficient Real-Time Emotion Detection Using Camera and Facial Landmarks- Binh T. Nguyen,Minh H. Trinh, Tan V. Phan and Hien D. Nguyen

Emotion recognition has been broadly investigated in both theory and applications in many research fields, for example, computer science, neuro science, biology, psychology and medicine. It is an important step for machine understanding of human behaviors. Typically, one possible way for building a human emotion recognition system is using facial landmarks, which can be considered as key points in human faces. It is interesting to note that each emotion has unique facial expression, especially on eyes, eyebrows and mouth. Therefore, facial landmarks can be recognized as crucial characteristics to describe human emotions. In order to detect emotions from human images, there were usually three following steps: face detection, feature selection from face images and emotion classification from the extracted features. To solve the problem, we firstly detect any face appearing in each frame. With each detected face, we extract and normalize the corresponding facial landmarks. Next, we calculate a set of features as inputs for emotion classification later. Finally, we choose an appropriate classifier by analyzing several techniques including SVMs.

Context in Human Emotion Perception for Automatic Affect Detection: A Survey of Audiovisual Databases Bernd Dudzik, Michel-Pierre Jansen, Franziska Burger

An important aspect of human emotion perception is the use of contextual information to understand others' feelings even in situations where their behavior is not very expressive or has an emotionally ambiguous meaning. For technology to successfully detect affect, it must mimic this human ability when analyzing audiovisual input. Databases upon which machine learning algorithms are trained should capture the context of social interactions as well as the behavior expressed in them. One of the goals in Affective Computing research is automatic affect detection – providing computers with a humanlike ability to perceive affective states in their users. Affect detection is primarily approached by annotating and automatically analyzing behavioral signals captured as audiovisual data. In doing so, most research to date has focused on interpreting these behavioral signals in isolation, while largely ignoring their surrounding context.

Chapter 3

Proposed System

We propose a system that will be able to detect a face from the input frame. Once the face is detected, using the convolutional neural networks from the facial characteristics the system will be able to recognise the emotions and then subsequently recognise age and gender as well.

Chapter 4

Technology Stack

- Python 3.7 - Python is an interpreted, high-level, general-purpose programming language. It will be used in this project for basic programming along with it's basic libraries
- Scikit Learn - Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.
- Google Colab - Colaboratory is a open source research tool for machine learning education and research. It's a Jupyter notebook environment that requires no setup to use and runs entirely on cloud.
- Tensorflow - It is an open source artificial intelligence library, using data flow graphs to build models. It allows developers to create large-scale neural networks with many layers.

Chapter 5

Project Design

5.1 Description of use case diagram

The user stands in front of the camera and looks into it. His/her face is detected from the input frame which is captured from the live stream. The system then subsequently detects emotion, age and gender and provides a possible result.

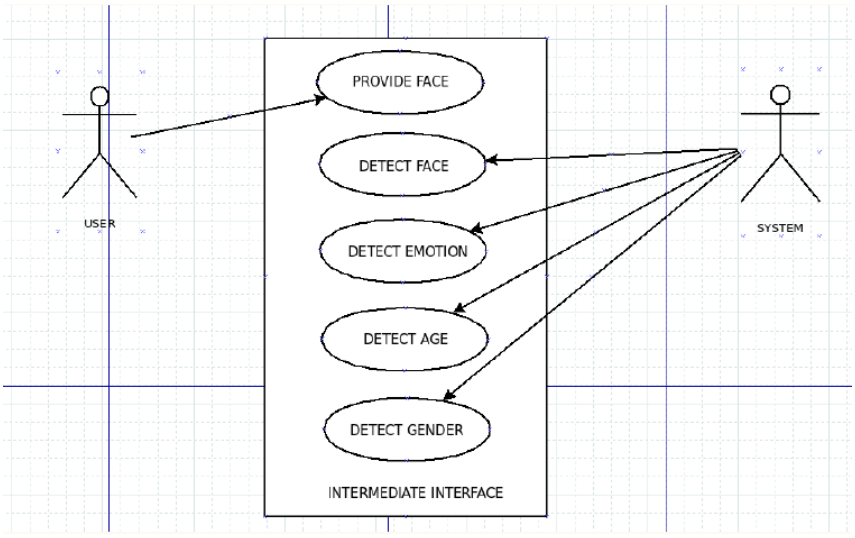

Figure 5.1: Use Case Diagram

5.2 Description of class diagram

Once the user provides his/her face, its detected through the face detection module. Every subsequent step is built on the result of the previous step or the initial step. The emotion recognition and age, gender recognition is built on face detection. Similarly the complete output is based on the result of the emotion and age,gender recognition modules separately.

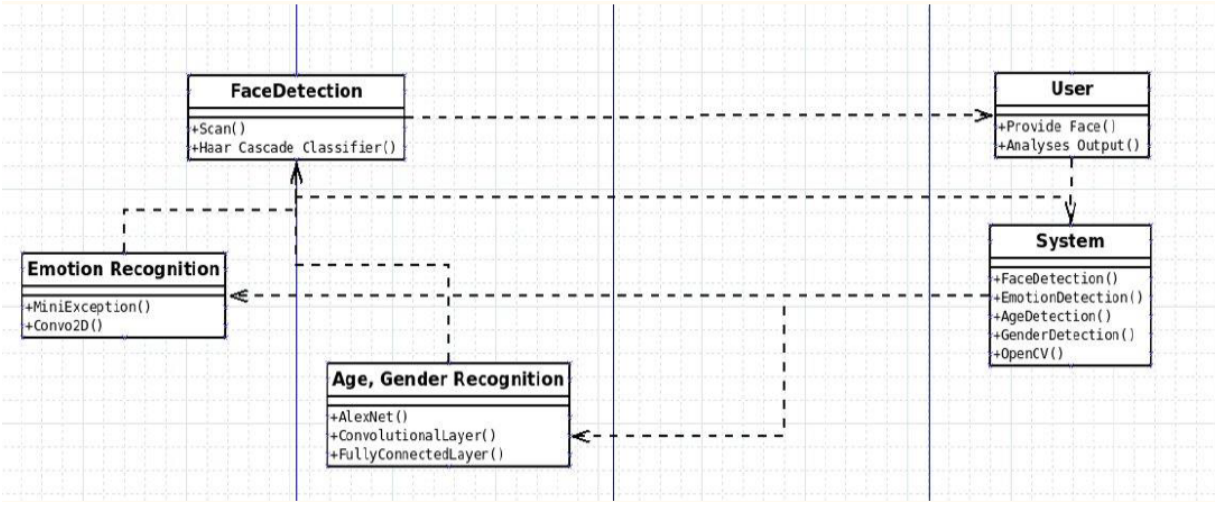

Figure 5.2: Class Diagram

Chapter 6

Flow of Modules

6.1 Face Detection

This module involves the detection of the face of the person standing in front of the camera. First, the photo is taken from the webcam stream live. Second, we turn the image to gray scale and use the Haar Cascade Classifier to detect faces in the image. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect faces in other images. Initially, the algorithm needs a lot of positive images (images of faces) and negative images (images without faces) to train the classifier. Then we need to extract features from it. For this, Haar features shown in below image are used. They are just like our convolutional kernel. Each feature is a single value obtained by subtracting sum of pixels under white rectangle from sum of pixels under black rectangle.The first feature selected seems to focus on the property that the region of the eyes is often darker than the region of the nose and cheeks. The second feature selected relies on the property that the eyes are darker than the bridge of the nose. Here features are applied step by step. If one window fails in a stage it is discarded and hence next stages are not calculated on it.

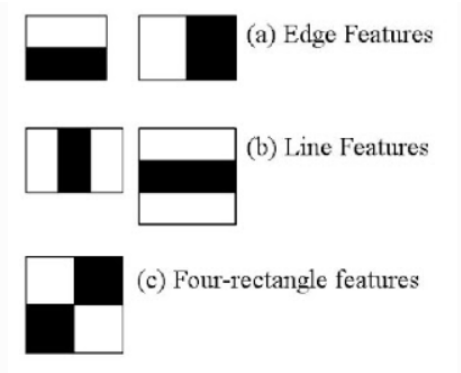

Figure 6.1: Features in Haar Cascade Classifier

6.2 Emotion Recognition

Once the face is detected, using the facial characteristics emotions are detected. Emotions can be classified into 7 different categories namely, happy, sad, anger, disgust, smile, fear and neutral. Each of these emotions has a unique set of characteristics that differentiates it from the rest of them. Successfully identifying these unique characteristics allows us to recognise the person's emotion effectively. We have used Mini Xception Model which uses Convolutional Neural Net is used to identify specific emotion out of the 7 listed emotions.

One significant advantage of deep networks is the automated learning of image representations, which are demonstrated to be more efficient than hand-crafted features. There are approaches which use pre-trained CNN's as feature extraction machines and combine the deep representations with classifiers such as linear SVM and multilayer neural network. In the performance of convolutional neural networks in large-scale video classification was studied, and it was found that CNN architectures are capable of learning powerful features from weakly-labeled data that far surpass feature based methods in performance and that these benefits are surprisingly robust to details of the connectivity of the architectures in time.

There are various techniques that can be kept in mind while building a deep neural network and is applicable in most of the computer vision problems. Below are few of those techniques which are used while training the CNN model below.

Data Augmentation : More data is generated using the training set by applying transformations. It is required if the training set is not sufficient enough to learn representation. The image data is generated by transforming the actual training images by rotation, crop, shifts, shear, zoom, flip, reflection, normalization etc.

Kernel-regularizer : It allows to apply penalties on layer parameters during optimization. These penalties are incorporated in the loss function that the network optimizes. Argument in convolution layer is nothing but L2 regularisation of the weights. This penalizes peaky weights and makes sure that all the inputs are considered.

BatchNormalization : It normalizes the activation of the previous layer at each batch, i.e. applies a transformation that maintains the mean activation close to 0 and the activation standard deviation close to 1. It addresses the problem of internal covariate shift. It also acts as a regularizer, in some cases eliminating the need for Dropout. It helps in speeding up the training process.

Global Average Pooling : It reduces each feature map into a scalar value by taking the average over all elements in the feature map. The average operation forces the network to extract global features from the input image.

Depthwise Separable Convolution : These convolutions are composed of two different layers: depth-wise convolutions and point-wise convolutions. Depth-wise separable convolutions reduces the computation with respect to the standard convolutions by reducing the number of parameters.
Following is the flow diagram depicting when each layer works in the architecture.

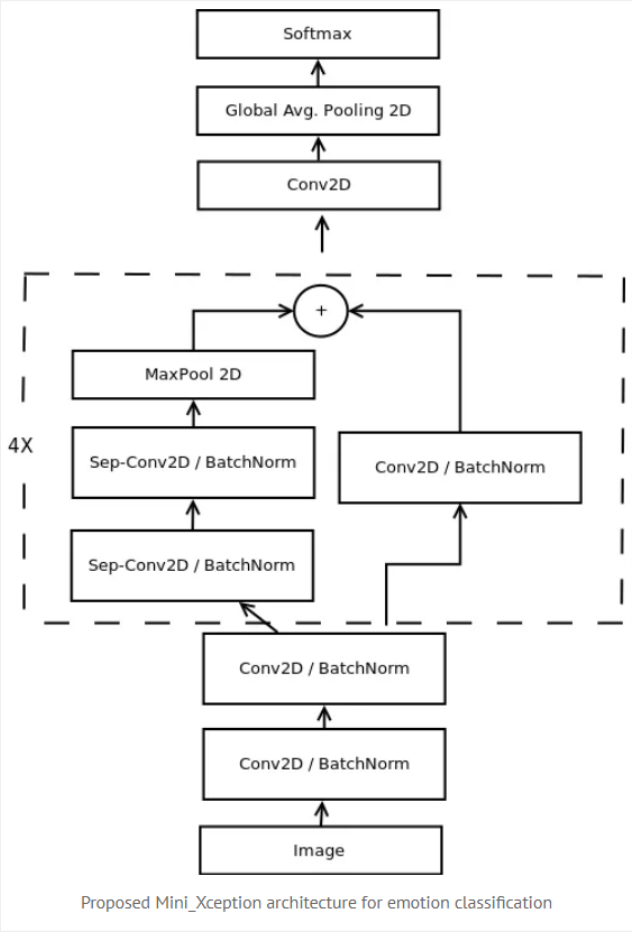

Figure 6.2: Mini Xception Architecture

6.3 Age and Gender Recognition

In machine learning, utilising the previously learnt knowledge for solving a new problem is called as knowledge transfer. The transfer learning using deep CNNs is very helpful for training a model with limited size dataset because CNNs are prone to overfitting with a small dataset.
We have used AlexNet Model which also uses Convolutional layers to recognise age and gender. We consider the age and gender as two separate classification tasks. First, we classify the whole dataset using the age classifier (older and younger). Then, we classify the whole dataset using the gender classifier (male and female). Each class is divided into two parts randomly: one part is used for training, while the other is used for testing the performance of the dataset. The input of AlexNet is 227 x 227 pixels RGB image and the network has 5 convolutional layers (from C1 to C5) and 3 fully connected layers (Fc6 to Fc8). This architecture contains 60 million parameters. Learning so many parameters for a few thousand training images of the new task is problematic and time-consuming; therefore, transfer learning is very suitable as a substitution.

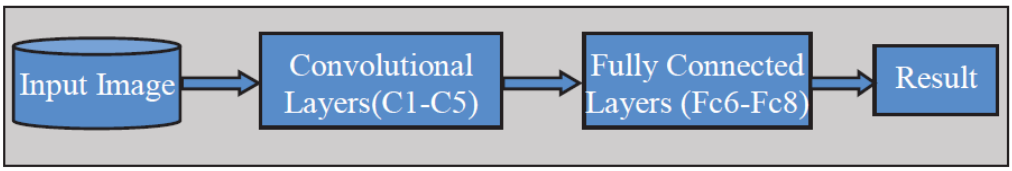

Figure 6.3: AlexNet Structure

There are a few characteristics that makes AlexNet different than the other architectures. They are as explained below.
Firstly, it does not use tanh function but uses ReLU function in the model.
Secondly, it allows multi-GPU training.
Thirdly, it icorporates data augmentation and dropout technique in order to avoid the over-fitting problem.

.

Chapter 7

Result

The output of the system is displayed at run time along with the live input stream. It consists of one window which shows what percentage of each of the emotion from the stated emotions is associated with the person's face in the input frame. Apart from that it also shows the age and gender of the person. Following are the examples of the same.

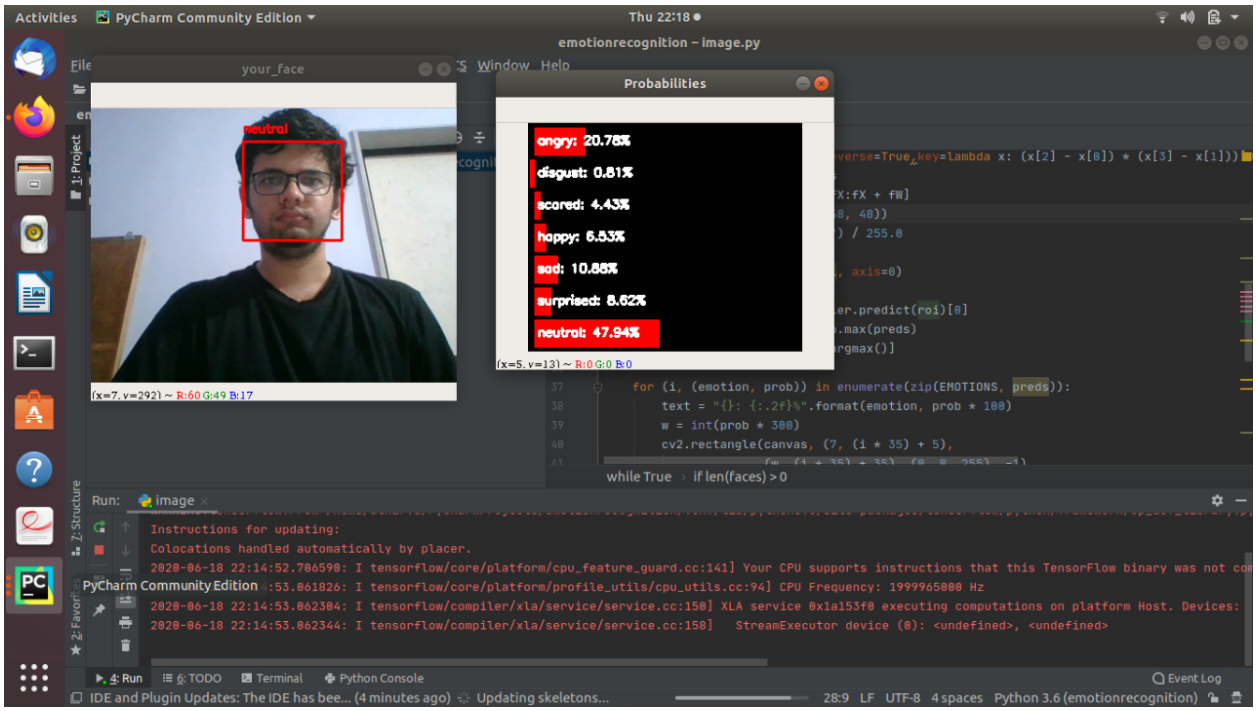

Figure 7.1: Percentage of each emotion that is associated with the image with Neutral being the highest

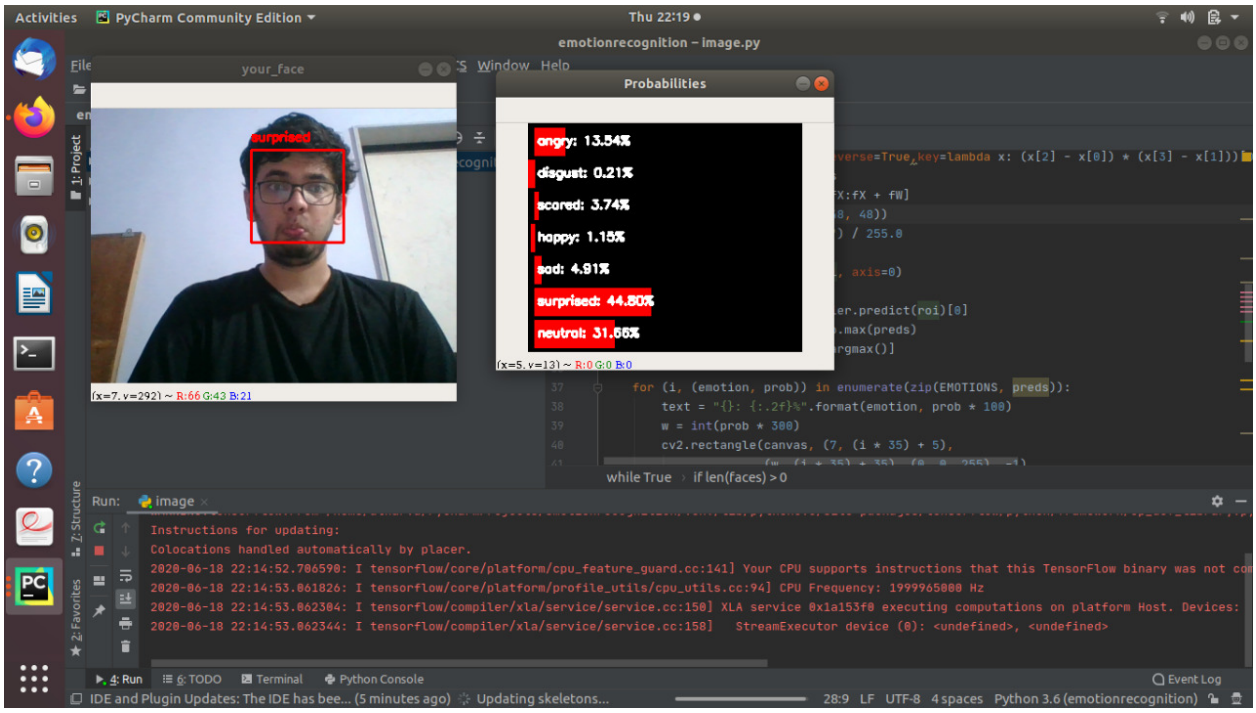

Figure 7.2: Percentage of each emotion that is associated with the image with Surprised being the highest

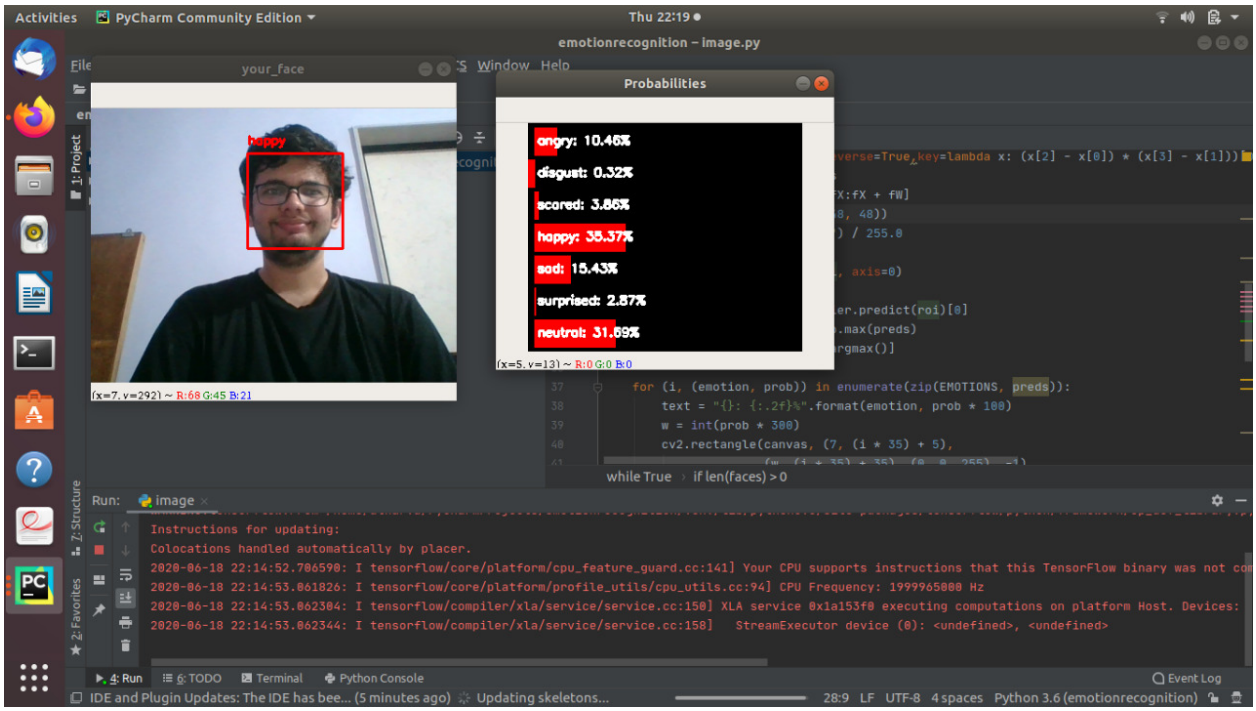

Figure 7.3: Percentage of each emotion that is associated with the image with Happy being the highest

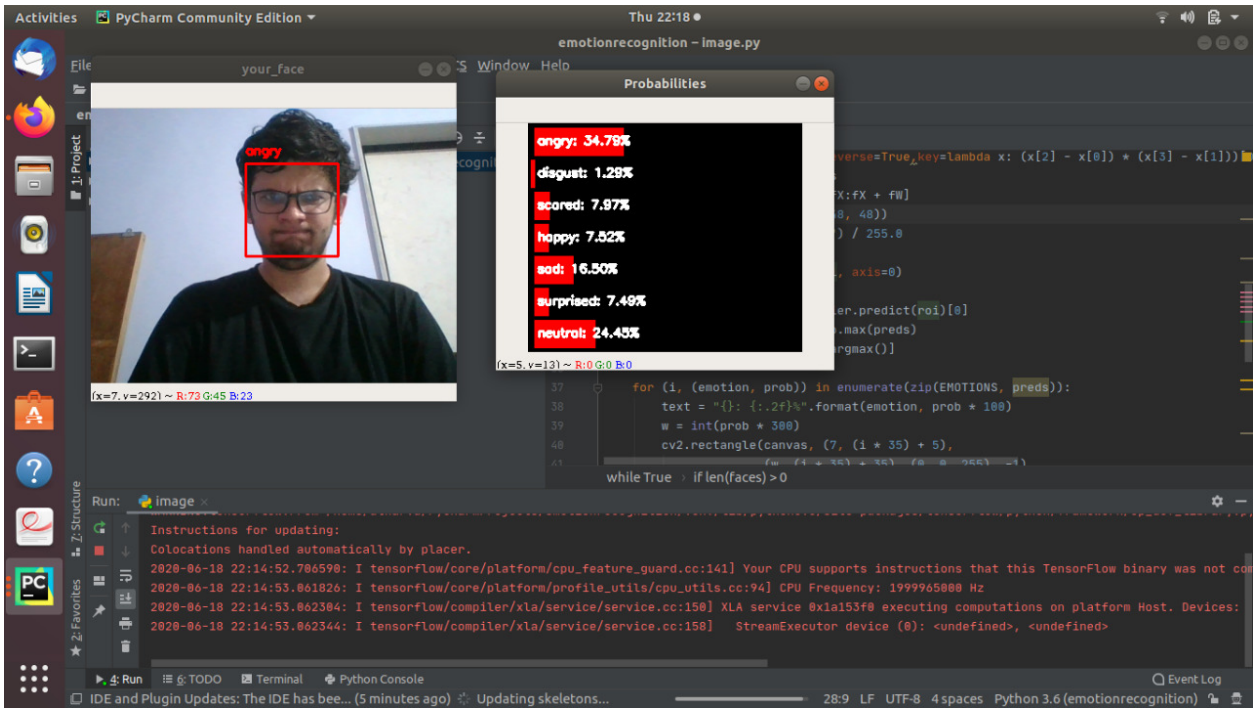

Figure 7.4: Percentage of each emotion that is associated with the image with Anger being the highest

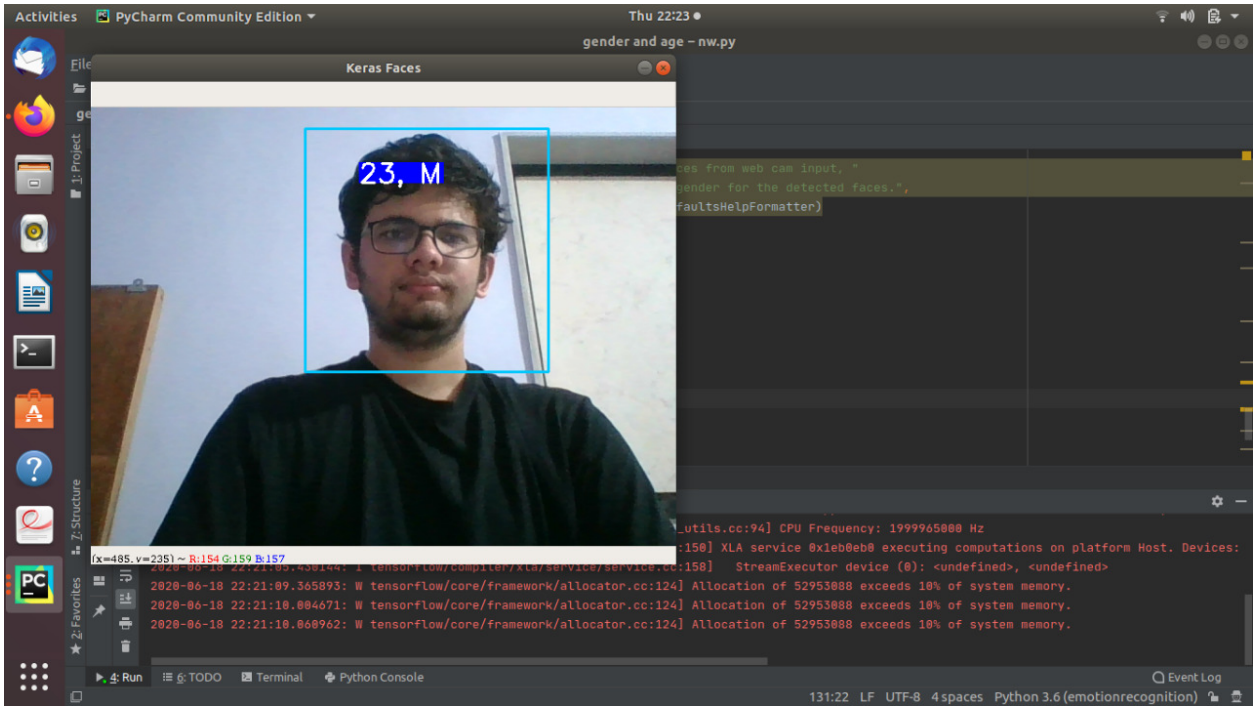

Figure 7.5: Age and gender of the person

Chapter 8

Future Scope

The proposed system is a technological advancement to the previous recognition or identifier systems such as bio-metrics.The system has a wide range of scope in fields of security and predictions. It enables us to detect and recognise a person faster and more efficiently. Hence finds a scope in the field of security. Also, the facial characteristics recognition along with the age and gender recognition will enable the said authorities to remotely detect and record the behaviour of a person and hence in some real time cases also predict the future actions based on the current characteristics.

Bibliography

[1] An efficient real-time emotion detection using camera and facial landmarks Binh T. Nguyen ; Minh H. Trinh ; Tan V. Phan ; Hien D. Nguyen 2017 Seventh International Conference on Information Science and Technology (ICIST) Year: 2017

[2] Context in Human Emotion Perception for Automatic Affect Detection: A Survey of Audiovisual Databases Bernd Dudzik ; Michel-Pierre Jansen ; Franziska Burger ; Frank Kaptein ; Joost Broekens ; Dirk K.J. Heylen ; Hayley Hung ; Mark A. Neerincx ; Khiet P. Truong 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII) Year: 2019

[3] Gender and Age Classification of Human Faces for Automatic Detection of Anomalous Human Behaviour Xiaofeng Wang ; Azliza Mohd Ali ; Plamen Angelov 2017 3rd IEEE International Conference on Cybernetics (CYBCONF) Year: 2017

[4] https://www.tensorflow.org/tutorials

[5] https://towardsdatascience.com

[6] https://www.dlology.com/blog/easy-real-time-gender-age-prediction

[7] https://medium.com/datadriveninvestor/l1-l2-regularization

Acknowledgement

I have great pleasure in presenting the report on **Human age, gender and emotion recognition** I take this opportunity to express my sincere thanks towards my guide **Prof. Archana Kotangale** Department of Computer Engineering, APSIT thane for providing the technical guidelines and suggestions regarding line of work. I would like to express my gratitude towards his constant encouragement, support and guidance through the development of project.

I thank **Prof. Sachin Malve** Head Deartment of Computer Engineering, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

I thank **Prof. Amol Kalugade** BE project co-ordinator, Deartment of Computer Engineering, APSIT for being encouraging throughout the course and for guidance.

I also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. I wish to express my deep gratitude towards all my colleagues of APSIT for their encouragement.

Student Name: Atharva Vaidya

Student ID: 16102043