EFFECTIVE TRACKING OF PASSING PEOPLE FOR MARKETING USING DEEP LEARNING

Submitted By

Aisha Rehman Burki	**01-132162-002**
Fatima Tuz Zahra	**01-132162-005**
Muhammad Ghazi Ud Din Zia	**01-132162-050**

Supervised by

Engr. Amna Waheed

Submitted to the department of computer engineering in the partial fulfillment of the requirements for the degree of Bachelors in Computer Engineering.

Department of Computer Engineering

Bahria University Islamabad.

2019

CERTIFICATE

Bahria University Islamabad
Department of Computer Engineering

Dated: <u>15-07-2020</u>

CERTIFICATE

We accept the work contained in the report titled "Effective Tracking of Passing People for Marketing using Deep Learning" as a confirmation to the required standard for the partial fulfillment the degree of Bachelor of Computer Engineering.

Mr. Waleed Manzoor
Project Coordinator
Date: **15-07-2020**

Engr. Amna Waheed
Supervisor
Date: **15-07-2020**

Engr. Ammar Ajmal
Internal Examiner
Date: **15-07-2020**

Dr. Usman Hashmi
External Examiner
Date: **15-07-2020**

Head of Department (CE)
Date: **15-07-2020**

UNDERTAKING

I certify that research work titled "*Effective Tracking of Passing People for Marketing using Deep Learning*" is my own work. The work has not been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged / referred.

Aisha Rehman Burki (47131)

Fatima Tuz Zahra (47134)

Muhammad Ghazi Ud Din Zia (47172)

DEDICATION

This thesis is dedicated to our parents for their love and support throughout the completion of our degree.

ACKNOWLEDGEMENTS

In the Name of Allah, the Merciful, the Beneficent. All praise be to Allah, the Lord of all worlds for bestowing us with endurance and strength to accomplish this project.

We would like to offer deep appreciation to our supervisor Engr. Amina Waheed for her expert knowledge, generous guidance, patience and encouragement throughout the completion of our project.

We are very much thankful to our parents for their untiring efforts and technical staff of Department of Computer Engineering for their generous support.

ABSTRACT

Today Pakistan’s local brands need to approach for new marketing schemes in order to attract new customers as well as to retain the existing ones so they can compete with their foreign competitors. Many retailing companies are gathering data of their customer’s on basis of age and gender but it is a manual process to do it. The solution that we are providing is basically based on hyper-targeting which is focused on individual group of people that will help retailers/marketers to have a complete knowledge of their customer’s demographics. The primary aim of this project is to determine in-store activity, help marketers analyze performance and success of their new product, manage staff schedules according to peak periods and maximize the sales potential. We are doing gender and age group segmentation with the help of deep learning algorithm i.e. Mini-Xception. In order to analyze the trends given from the statistical information through our system, customer’s information is displayed on a dynamic web application. The end application is a web service on which accurate demographics of the customers will be shown.

Keywords:

Mini-Xception, Deep Learning, Marketing

TABLE OF CONTENTS

LIST OF FIGURES

Number *Page*

LIST OF TABLES

CHAPTER 1

Introduction

Today Pakistan's local brands need to approach for new marketing schemes in order to attract new customers as well as to retain the existing ones so they can compete with their foreign competitors. Many retailing companies are gathering data of their customer's on basis of age and gender but it is a manual process to do it. The solution that we are providing is basically based on hyper-targeting which is focused on individual group of people that will help retailers/marketers to have a complete knowledge of their customer's demographics. The primary aim of this project is to determine in-store activity, help marketers analyze performance and success of their new product, manage staff schedules according to peak periods and maximize the sales potential. We are doing gender and age group segmentation with the help of deep learning algorithm i.e. Mini-Xception. In order to analyze the trends given from the statistical information through our system, customer's information is displayed on a dynamic web application. The end application is a web service on which accurate demographics of the customers will be shown.

1.1 Overview

Marketing strategies cannot be formed without knowing your audience towards a specific product or brand. In today's world of marketing customers demographics play very important role. To help marketers analyze a complete record of their customer's demographics we are proposing a solution which will accurately identify the age group and gender of the consumers.

1.1.1 System Level Diagram

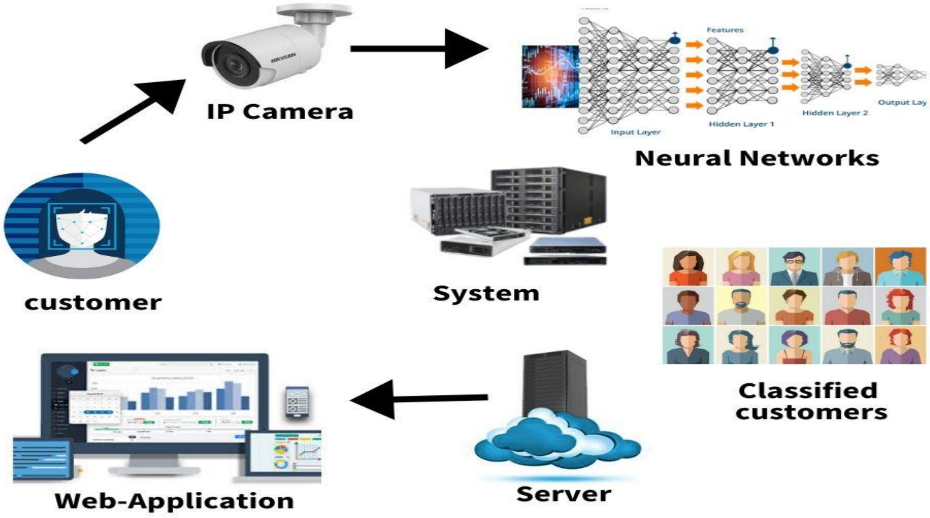

Fig 1.1: System Level Diagram of Project

Figure 1.1 depicts the system level diagram of the project. As soon as the customer will enter the store, the IP camera will capture its presence. Then from the IP camera frames will go into our system which is trained on CNN based model and it will classify the customers w.r.t age and gender. The customer count w.r.t age and gender will be stored in database. The data base is connected to web application, where exact customer demographics will be displayed on the dashboard.

1.1.2 Artificial Intelligence

The ability of a computer program to make a machine or system think, learn and act like humans is known as artificial intelligence. It can be described as an ability of a machine to mimic as human functions. It is also referred as system intelligence. If we look around our surroundings we will notice that from our cellphone's camera to self-driving cars we have artificial intelligence everywhere growing rapidly. Research in artificial intelligence aims to make systems reason; help systems make rational decisions, process natural language and the ability to manipulate. Artificial intelligence is found in applications i.e. medical diagnosis, computer search engines, and handwriting recognition as well as computer vision. AI systems are embedded in numerous systems as well. Some of the major activities that computers are designed to do include:

1. Learning
2. Computer vision
3. Speech recognition
4. Natural language processing
5. Complex problem solving

1.1.3 Machine Learning and Deep Learning

Artificial intelligence is a wider concept whereas machine learning is a subset of artificial intelligence. It is basically one of the most common applications of artificial intelligence. Machine learning involves the study of algorithms that machine uses to perform complex, specific tasks without using explicit instructions by the programmers. To develop computer programs that can access data and use it to learn from them is the focal point of machine learning.

Deep learning is a subset of machine learning. Machine learning is used when we have less amount of data and all the features are known where as in deep learning we deal with large amount of data especially pictures and all the features are not manually fed into the machine, the deep learning algorithms extracts the features themselves.

We chose this field because of the characteristics mentioned above. In our project we have taken the help of this field to propose a solution that will help the retailing sector in managing their brands and marketing processes.

1.1.4 Automated Statistical Analysis

The need of automation is increasing rapidly and Pakistan's retailing sector faces a lot of issues in manually doing the marketing process. As Pakistan's retailing sector is growing immensely, various types of retailing companies are gathering data of their customers but manually. To gather customer segmentation is a very tedious task and it often leads to customers despise towards the product/brand. Therefore we are providing a solution which caters all these problems and shows statistical analysis of customers automatically with the help of facial detection.

1.2 Problem Statement

The problem statement of our project is:

With the increase of global brands and outlets, Pakistan's retail sector at present is witnessing an impressive growth and local retailers are competing with their larger foreign competitors. This system will provide a complete marketing workout i.e. it will provide a specific demographic of customers in real time using video analysis.

1.3 Motivation

The need of automation is increasing rapidly and Pakistan's retailing sector faces a lot of issues in manually doing the marketing process. In industrial automation the recent automation is based on quality and flexibility.

As Pakistan's retailing sector is growing immensely, various types of retailing companies are gathering data of their customers but manually. To gather customer segmentation is a very tedious task and it often leads to customers despise towards the product/brand.

The solution that we are providing is basically based on hyper-targeting which is focused on individual group of people that will help retailers/marketers to have a complete knowledge of their customer's demographics. The primary aim of this project is to determine in-store activity, help marketers analyze performance and success of their new product, manage staff schedules according to peak periods and maximize the sales potential.

1.4 Objectives

The objectives of our project are as follow:

- To classify gender and age group of customers based on facial features.
- To design a stand-alone software for local stores to target their customers.
- To analyze and identify accurate customer's demographics with peak periods.

- To help marketers maximize sales potential and analyze the success of marketing initiative.

1.5 Thesis Structure

Chapter 2 provides precise information about our literature review and the algorithms which we have used, it includes initial research work to the latest research work in this field with respect to the best algorithm that we chose for our project. In Chapter 3 detailed discussion is made about the methodologies of our module along with web application. Chapter 4 consists of all the experimentations and results. Chapter 5 provides conclusion and further work is mentioned in Chapter 6.

CHAPTER 2

Literature Review

2.1 Neural Networks

Neural Networks form the base of deep learning, a sub field of machine learning where algorithms are inspired by the structure of the human brain. Neural networks take data as input and train themselves to recognize the patterns in the input data, and then predict the output for the new set of similar data. Neural networks assist us in clustering and classification of the data. We can consider them as a classifying layer on the top of our data.

Neural Networks are made up of layers of neurons which are the core processing units of a network. At the beginning we have the input layer which receives the input and at the end we have output layer which predicts our final output. In between these two layers we have multiple hidden layers which perform most of the computation required by our network.

Learning from data automatically seems promising. In the past years neural networks were only used to overcome specific issues but with the passage of time new techniques were generated to overcome more complex issues. Those new techniques are now termed as deep learning. With further development in deep learning techniques and deep neural networks very astonishing results are achieved in computer vision, speech recognition, and natural language processing on many significant problems. Companies such as Google, Microsoft and Facebook use them on a large scale for automation and real time decisions.

A simplest neural network is shown in figure below:

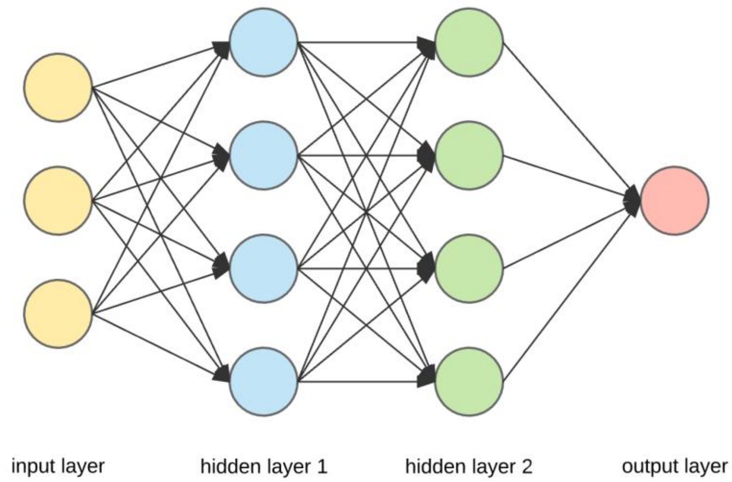

Fig 2.1**:** Simple architecture of Neural Network

Figure 2.1 represents the simple architecture of neural networks. Neural networks basically comprises three main layers i.e. input layer, output layer and hidden layer, there can be multiple hidden layers. Each layer comprises of one more nodes and each node is connected to all the other nodes in the next layer. The flow of information is indicated by the lines between the nodes.

2.1.1 Convolutional Neural Networks (CNN):

A convolutional neural network also known as CNN, it is an artificial neural network that has been widely used for image analyzation. CNN is now the go-to model for every problem related to the image. We can think of CNN as artificial neural network which has capability of taking an image as an input and then it can pick out or detect patterns and make sense out of them. CNN basically assigns learnable weights to several features present in an image enabling it to differentiate them from each other. This pattern detection is what makes CNN very effective is image analyzation. The plus point of using CNN is that the preprocessing required for it is comparatively very low as compared to the algorithms of classification.

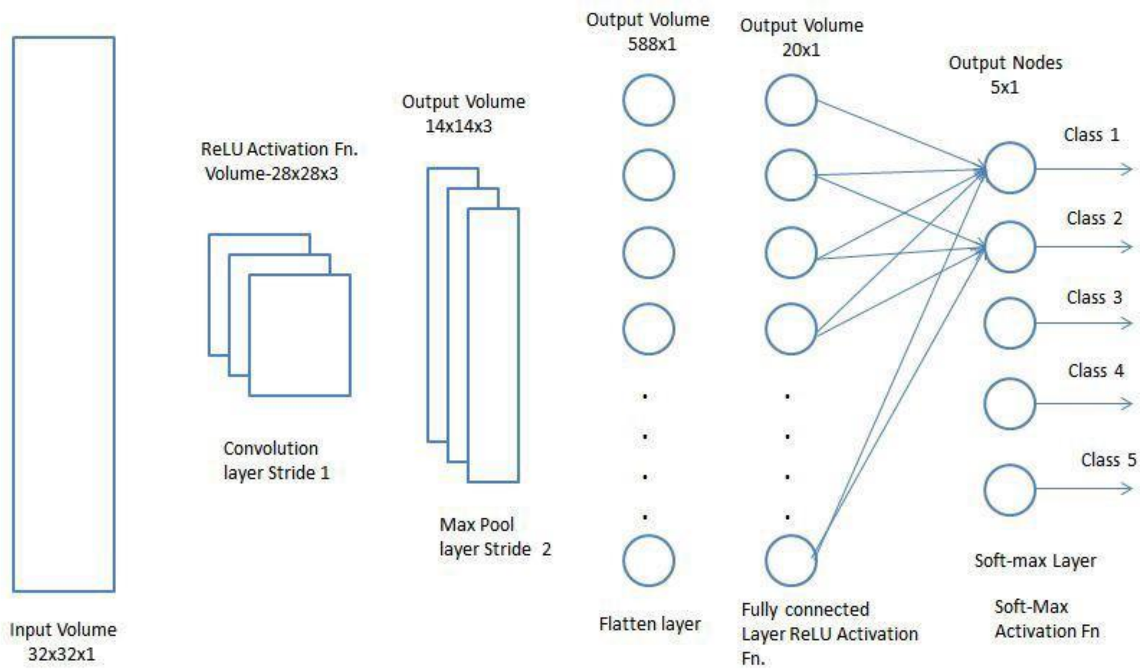

Fig 2.2**:** Simple architecture of CNN [1]

Figure 2.2 represents simple architecture of CNN i.e. it consist of input layer, convolutional layer, pooling (either max pooling or average pooling), a flatten layer or dense layer, fully connected layer where we apply Rectified Linear Unit (ReLU) and softmax layer.

2.2 Gender Classification using Facial Features

As we are moving into the age of automation the classification of gender by using facial features with higher accuracy is the utmost desire of everyone. But still a lot of room for improvement is still left for the performance of already proposed methods on real world images. There are a lot of numbers of algorithms developed for this task. The two main algorithms that we used in our model are explained below in detail.

2.2.1 Mini-Xception

Mini-Xception is mostly based on the schemes of previous architectures but Ariaga, Ploger and Valdenego [2] proposed an architecture in which the last fully connected layer is being removed by using global average pooling. The last layer in this architecture had the exact same amount of features as the number of classes for the classification. Suggested architecture is a standard fully-convolutional neural network which consists of 14 convolution layers, ReLUs, Batch Normalization and Global Average Pooling.

This model combines the sequential-fully CNN and the Xception model.

- Sequential-fully CNN:

  This architecture uses Global Average Pooling to completely remove any fully connected layers.

- Xception:

  This architecture combines the use of residual modules and depth-wise separable convolutional network.

For the training of this model IMDB dataset was used which consist of 460,724 RGB images and which are divided into two classes i.e. men and women.

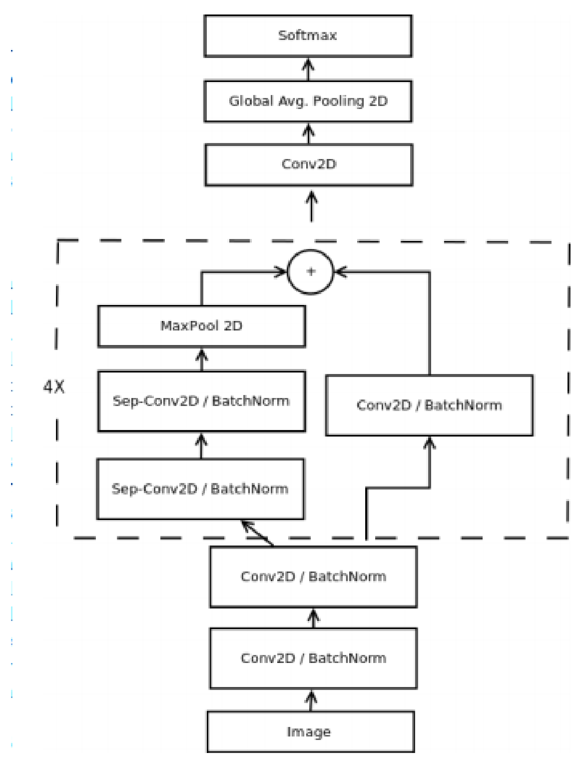

Fig 2.3**:** Proposed architecture of Mini-Xception [2]

Figure 2.3 represents the architecture of Mini-Xception model which is basically comprised of sequential fully connected CNN and Xception model. Xception model is within the doted boundary, it further consist Resents and depth wise separable convolutional network. In Mini-Xception model the Xception layers are repeated four times. Mini-Xception model takes an input of size 64x64.

The results that were extracted by using the above model are given below. In fig 2.4 the colour blue represents women class and the color red is assigned to men class.

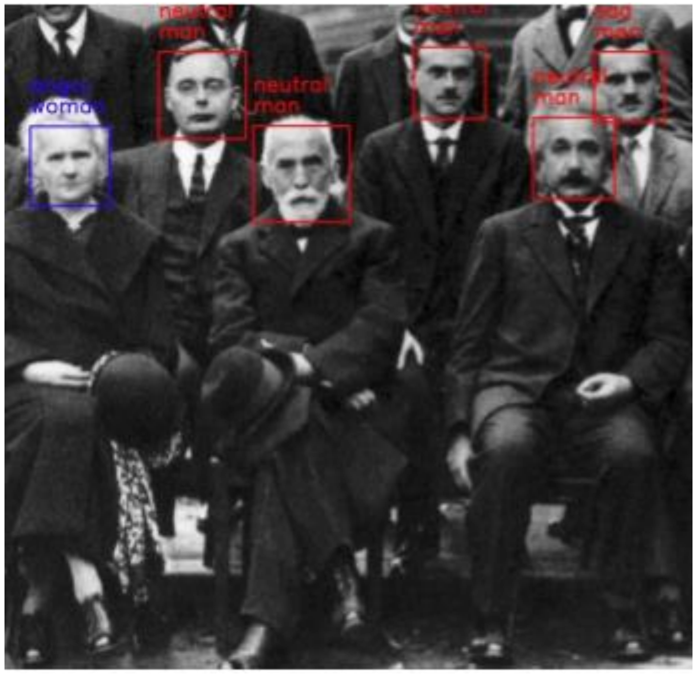

Fig 2.4: Results of Mini-Xception [2]

2.2.2 Densely Connected Neural Networks

Research in past has proved that Neural Networks can be more accurate and effective if the connections in the layers closer to input and output layer are shortened. Gao, Zhuang and Laurens [3] Dense Net is one of the convolutional neural networks. Its architecture is given below:

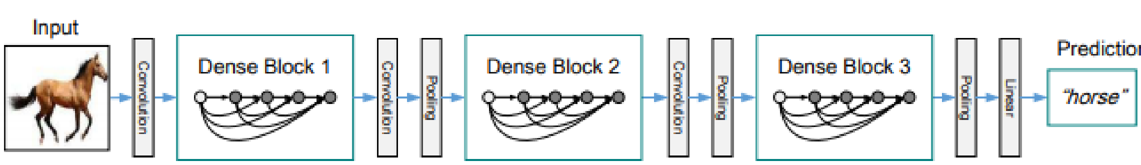

Fig 2.5: A Dense Net with three dense blocks comprising of transition, pooling and convolutional layer. [3]

In Dense Net all the matching features are connected to the layers allowing a layer to have the access to all the preceding features within its dense block while a traditional neural network only uses the most recent features. Dense Net at each layer implements a nonlinear operation which can be the joint operation of functions such as Batch Normalization (BN), rectified linear units (ReLU), Pooling, or convolution (Conv).

According to Gao, Zhuang, Laurens and Kilian Q. [3] pooling and average pooling was done to squeeze the image and classification of data in specific image respectively. The initial convolution layer comprises 2k convolutions of size 7x7 with stride 2; the numbers of feature-maps in all other layers also follow from setting k.

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

Figure 2.6: Growth rate for all DenseNet networks is k=32

Figure 2.6 represent different configuration of Dense Net architectures i.e. DenseNet-121 consist of 121 layers. Each Dense block consist of connected layers; each block is then followed by a transition layer which consists of convolution and pooling layer. Growth rate defines the number of channels added in each layer.

2.3 Algorithm for Face Detection

Face detection is a tricky task because while detecting a face there are certain factors that should be kept in mind such as occlusion, focus, light and pose etc. The face detection is done on the basis of features, appearance and by template matching.

2.3.1 Haar Feature-based Cascade Classifier

Haar features were proposed by Hungarian scientist Alfred Haar in 1909. We assign haar features to every single important feature of human face. Edge features can detect edges and line features can detect lines quite effectively.

Haar cascade classifier detects faces on the basis of Haar features i.e. edges and lines. This Haar cascade algorithm is trained on positive and negative samples. Positive samples are images that contain faces whereas negative samples are images which do not contain faces.

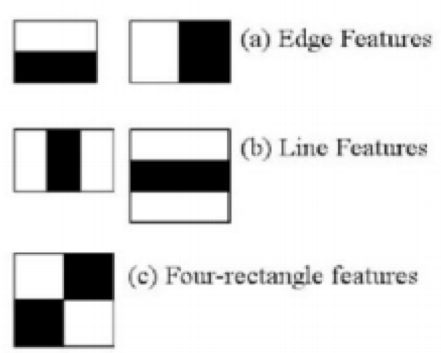

Fig 2.7: Representation of Haar Features which Haar cascade classifier uses to detect faces. [4]

While training each of the above mentioned features is applied to the training images in order to find the best face and non-face regions. If the image contains face regions then we keep it otherwise we discard the image. By using Haar cascade classifier a lot of time is saved as we only have to deal with images that contain faces.

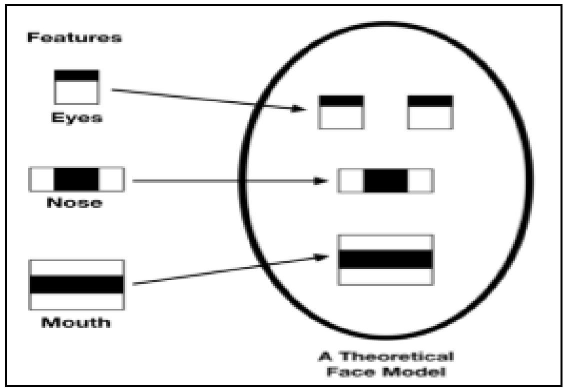

Figure 2.8: Feature Identification [4]

Haar-cascade_frontalface_default.xml file is used for feature extracting process.

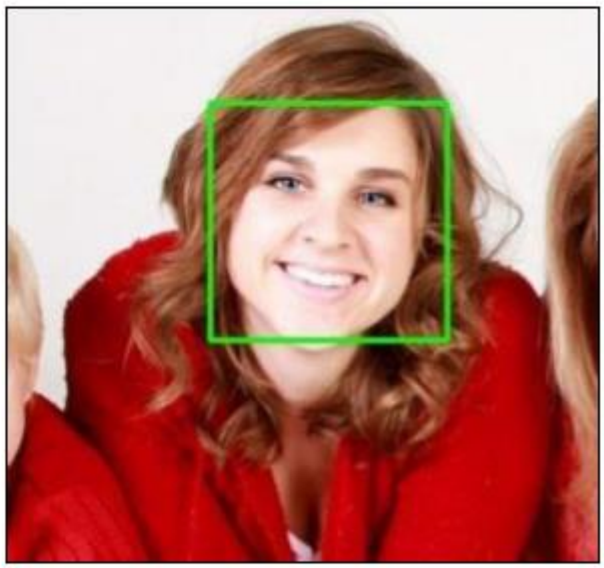
Figure 2.9: Haar Cascade Frontal Face Detection [4]

2.3.2 Dlib Face Detector

Dlib is a C++ library/toolkit that contains machine learning algorithms, including computer vision. There are two types of Dlib face detector. One is based upon HOG and SVM and the other one is CNN based Dlib face detector.

Dlib based upon HOG has faster CPU speed and shows great results in different scenarios. Either it is a frontal face or non-frontal it works quite effectively. It consists of 2825 training images. It is the more reliable and faster method on CPU

CNN based Dlib detector uses Maximum Margin Object Detector (MMOD). It consists of 7220 training images and it works well on GPU only. It performs really well under different face angles and occlusions.

The results of both Dlib detectors are given below.

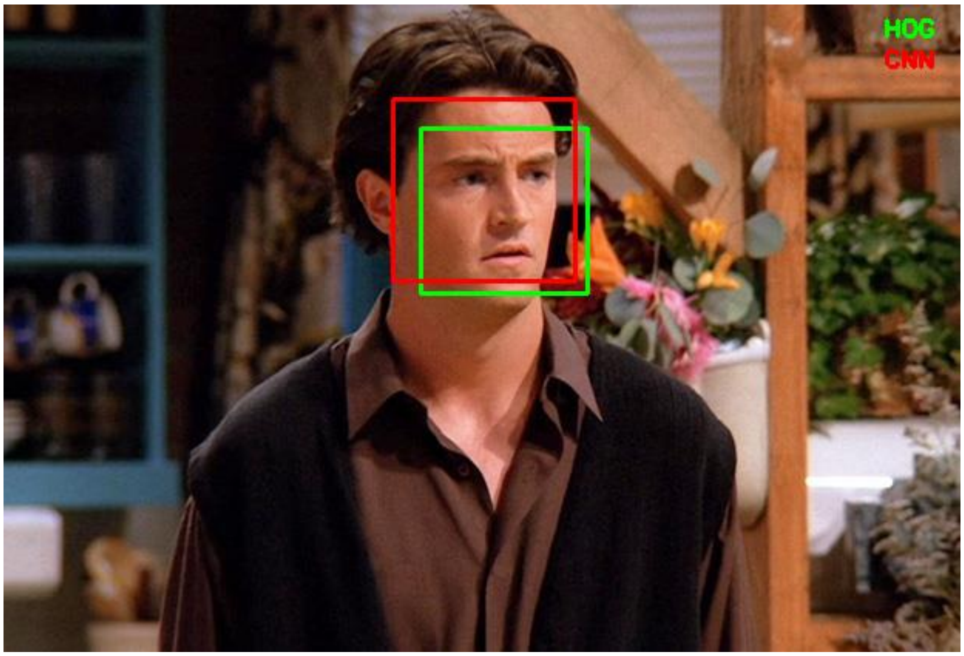

Figure 2.10: Dlib HOG and CNN frontal [6]

Figure 2.11: Dlib HOG and CNN non-frontal [6]

From the above results (figure 2.10 and figure 2.11) we can clearly see that CNN based Dlib is more effective as it works accurately in detecting non-frontal faces. But the one and only drawback that it has is that it requires a lot amount of computational power.

2.3.3 Deep Neural Network Face Detector

Out of all the face detecting methods discussed earlier DNN is the best method. It works with greater accuracy and precision under different frontal angles and even with occlusion. It is just slower than the Haar based detector else it has edge over it in every way.

2.4 Comparison of Algorithms

In order to decide which algorithm is better and must be opted they were tested under different factors and circumstances which are explained in detail below.

2.4.1 Comparison on Different Angles

We tested the algorithms on different angles and the result of all algorithms is given below.

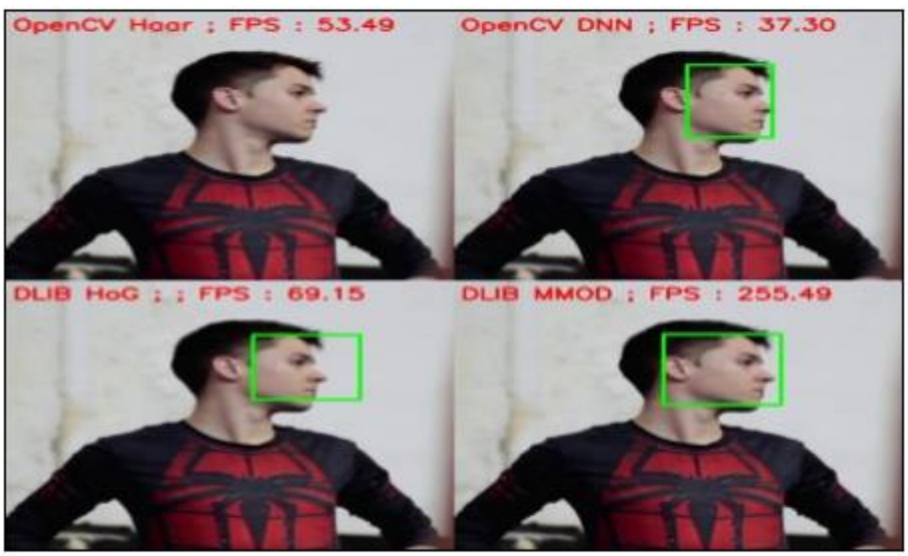

Figure 2.12: Comparison on Non-Frontal faces [6]

2.4.2 Occlusion

Occlusion is the condition where there is a hindrance in detecting the face. The hindrance could be of any type whether it is a cap, scarf, hand or a face hidden behind a face etc. It is clearly the most challenging aspect of face detection. Given below are the results showing how face detecting algorithms work under occlusion.

Figure 2.13: Comparison on Occlusion [6]

2.4.3 Speed

The speed comparison of algorithms is given below.

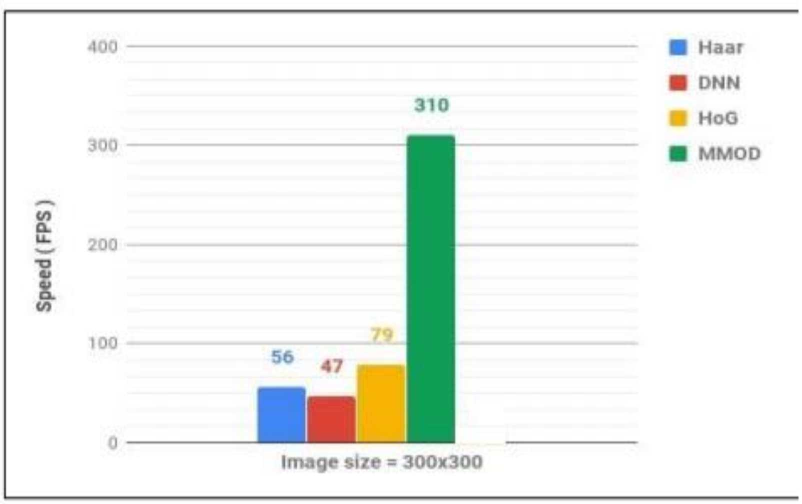

Figure 2.14: Comparison on basis of speed

2.5 Video Analytics in Marketing

Tetyana Kholod, Sule Balkan [7] did a research on video analytics that whether it could prove to be a game changer in the marketing world or not. According to their research it was shown that video analytics could be one of the most important elements in the marketing world. To get manual feedback from customers is a very tedious task as it involves the process of performing proper analysis. Video analytics requires zero response from the customer, thus making it very effective. The video analytics imitates the observation method of market surveys without the observer's biasness and providing more reliable and accurate data.

Table 2.1: Literature review of gender classification, face detection and video analytics

Sr. No.	Authors	Year	Techniques Used	Database	Results
1.	Octavio Arriaga, Matias Valdenegro-Toro, Paul Plöge [2]	2017	CNN, Mini-Xception	IMDB-Wiki dataset, FER-2013 dataset	Accuracy 96%
2.	Ke Zhang, Ce Gao, Liru Guo, Miao Sun,Xingfang Yuan,Tony X. Han,Zhenbing Zhao, Baogang Li [8]	2017	Residual Networks of Residual Networks (RoR)	Adience dataset, ImageNet dataset, IMDB-Wiki dataset	Accuracy 93.24%
3.	Grigory Antipov , Moez Baccouche , Sid-Ahmed Berrani, Jean-Luc Dugelay	2016	Deep Learning Models	IMDB-Wiki dataset, Manually collected Children	Error=0.26

	[9]			Images dataset	
4.	Zhiqin Zhang [10]	2016	Convolutional neural network using the GoogLeNet architecture	IMDB-WIKI dataset	—
5.	Rasmus Rothe, Radu Timofte, Luc Van Goo [11]	2015	Deep Expectation, CNN VGG-16 architecture	ImageNet dataset, IMDB-Wiki dataset	—
6.	Philip Smith, Cuixian Chen [12]	2018	VGG19 and VGGFace models	Morph II dataset	Accuracy 95.7%
7.	Afshin Dehghan Enrique G. Ortiz Guang Shu Syed Zain Masood [13]	2017	Deep convolutional neural networks	Self-collected dataset	Accuracy 91%
8.	Xuan Liu, Junbao Li, Cong Hu, Jeng-Shyang Pan [14]	2017	Deep Convolutional Neural Networks using GoogLeNet	Adience dataset	Accuracy 90.37%
9.	Bartłomiej Hebda ; Tomasz Kryjak [15]	2016	Compact deep convolutional neural network	FERET and the Adience	98.6% accuracy for gender and

			(DCNN) architecture	Benchmark databases	86.4% for age
10.	Vishwanath A. Sindagi , Vishal M. Patel [16]	2015	Multi-task cascaded CNN	shaghaitec dataset	Mean Absolute Error (MAE) 20.0
11.	Manik Sharma, J Anuradha, H KManne and G S CKashyap [5]	2017	Opencv and haarcascade classifier	Self-Collected	Limitations: Cannot detect faces with eye wear
12.	Valerio Nogueira , Hugo Oliveria, Jose Aujusto Silva, Thales Vieira , Krerley Oliveria [17]	2019	CNN regression model	Self-collected	Error: 10.78%
13.	Tetyana Kholod, Sule Balkan [7]	2014	Video Analytics	—	Video analysis are better than manual surveys as no response from audience is required
14.	T. Revathi , T. M. Rajalaxmi	2019	Deep Convolutional	WIDER face	90.3%

	[18]		Neural Networks	dataset, FDDB dataset	
15.	Davis E.King [19]	2015	MMOD(Maximum margin object detector)	Self Collected	—

CHAPTER 3

Design and Methodology

3.1 Workflow

The very basic idea of the system is to classify gender and age group in real time. The workflow is illustrated in the Figure 3.1 given below. The figure shows how first live feed is being obtained from an IP Camera. This live feed is then given to the system in the form of frames and pre-processing techniques such as grayscale conversion are performed. For the extraction of human faces from the incoming frames, Dlib face detection is being used. The extracted faces are the region of interest and are being used for testing the results of our system. The faces are passed to the convolutional neural network, which have been trained using the IMDB-WIKI [20] dataset along with our local dataset, to classify them into their respective gender and age groups. Each time a certain gender or age group is detected; they are also counted. These results are stored in a database and displayed using a web application that is made available to the user.

3.1.1 Flowchart

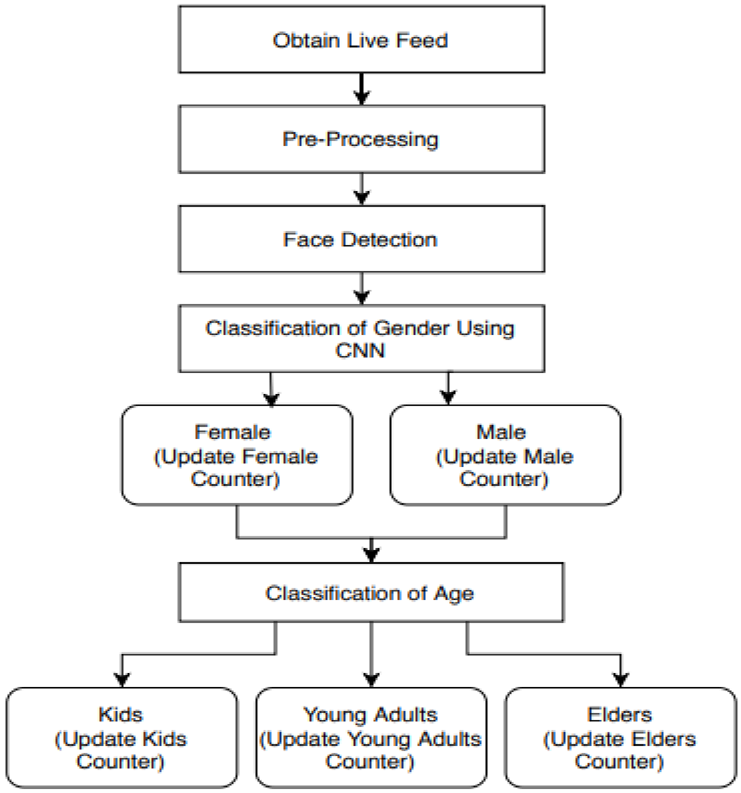

Figure 3.1: Flowchart

3.2 System Design

The hardware portion of this system consists of an IP Camera. As discussed, the Camera is simply being used to obtain live feed, which is fed into the system in the form of frames for gender and age classification. The software portion of the system is a web application which displays the data collected. Its design and working is discussed in the following chapter.

3.2.1 IP Camera Configuration

The camera being used is an 5-megapixel camera from HikVision. The configuration of the camera can be done by following these simple steps:

i. The first step is to install SADP Tool.

ii. After connecting IP camera to computer, check camera's IP using SADP tool.

iii. Make sure that both IP camera and computer have the same subnet mask.

iv. If the camera’s subnet mask is different from that of computer’s subnet mask, use the SADP tool to change Internet Protocol Version 4 (TCP/IPv4) of the camera. The camera should be connected to the computer after this.

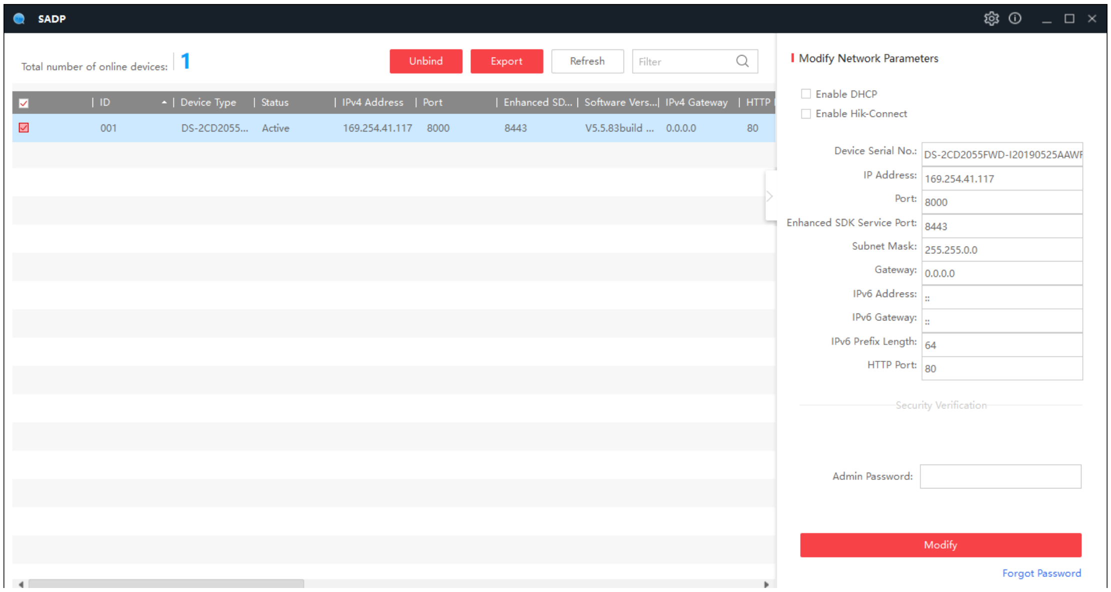

Figure 3.2: Setting up IP for Camera

The Figure 3.2 given above shows the SADP tool. This tool can be used to view the IP for the camera, the http port to view live stream, the serial number for the camera as well as its subnet mask. The IP, along with other parameters of the camera can be changed according to the user using the SADP tool.

3.3 Face Detection Tools

We tested two face detection models, OpenCV and Dlib. Dlib is further of two types, one is based on HOG+SVM and the other on CNN. After comparison between OpenCV and Dlib, based on speed, accuracy and certain pros and cons, we have selected Dlib for use. Following images are some of the results achieved by using Dlib.

Figure 3.3: Face Detection using Dlib

Figure 3.3 shows the results of using Dlib on an image containing focused frontal faces.

Figure 3.4: Face Detection on Non-Frontal Faces using Dlib

Figure 3.4 shows the results of using Dlib on an image containing non-frontal faces.

Figure 3.5: Face Detection on Faces with hat and eyewear using Dlib

Figure 3.6: Face Detection on Faces with eyewear using Dlib

Figure 3.5 and 3.6 shows the results of using Dlib on images containing eyewear, hats and different orientation of faces.

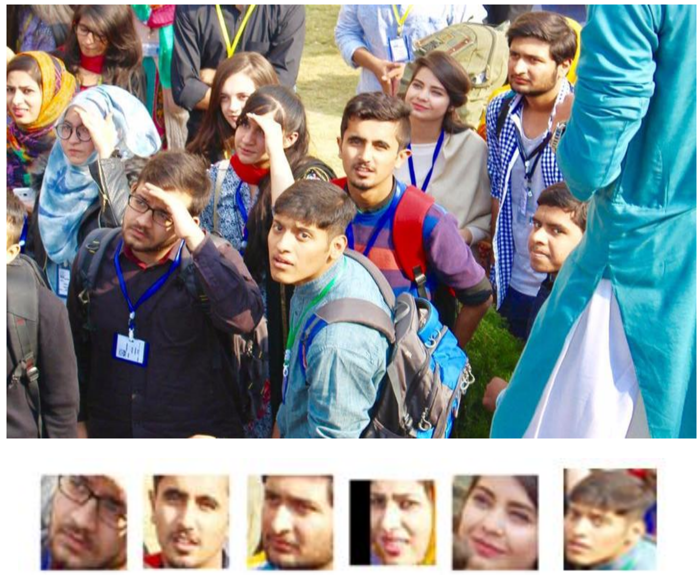

Figure 3.7: Face Detection under different lighting conditions using Dlib

As can be seen from the image in Figure 3.7, Dlib detects faces with bright and dull lighting as well as faces covered with hand.

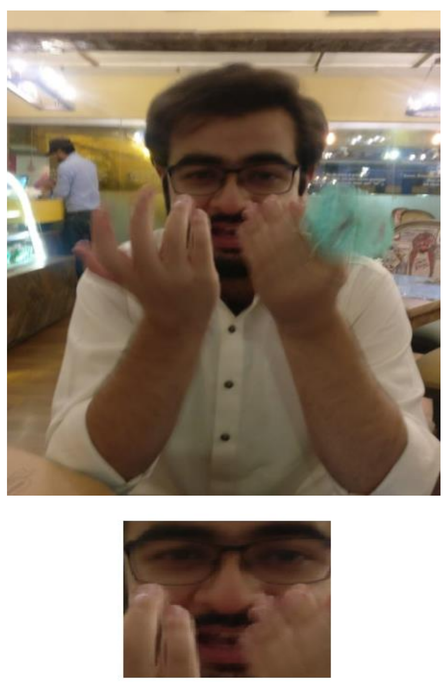

Figure 3.8: Face Detection on blurred image using Dlib

A face that is being covered by hands can be detected with Dlib. This can be seen in Figure 3.8. Blurred faces can also be detected.

3.3.1 Face Detection Comparison

Table 3.1: OpenCV and Dlib Comparison

Sr. No.	OPEN-CV	DLIB
1.	OpenCV is a C/C++ library of functions dealing with real-time computer vision.	DLib is a C++ library that contains machine learning algorithms, including computer vision.
2.	The major drawback of this method is that it gives a lot of false predictions.	More accurate results as compared to Open–CV.
3.	Doesn't work on non-frontal images.	Works for different face orientations.
4.	Doesn't work under occlusion.	Robust to occlusion.
5.	Very slow on GPU.	Works very fast on GPU.
6.	Open-CV uses Haar Cascade algorithm.	Dlib uses two types of feature detection algorithms i.e. HOG ,CNN.

Dlib gives better performance while detecting non-frontal faces and also for occlusion. For this reason Dlib is being used as the main tool for face detection.

3.3.2 Extracting Region of Interest

The region of interest refers to that part of the data, that needs to be identified for a particular reason. For this system, we need to be able to correctly detect human faces from the incoming images. If we use a face detection tool that has poor accuracy it will slow down processing and give inaccurate results during the classification process. Figure 3.9 given below, shows how a face is first detected in an image and then certain features within the face are extracted.

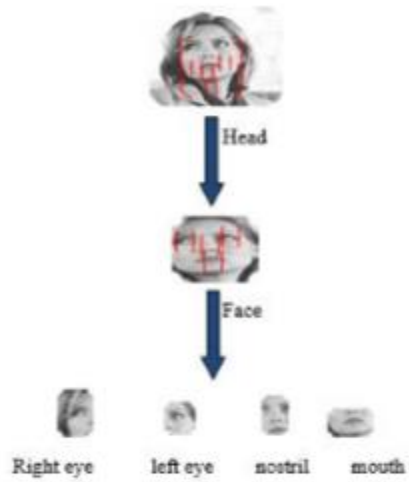

Figure 3.9: Extracted features from a frame

3.3.3 Real Time Face Detection through Live Feed

Face detection on static images is different from face detection in real time. In real time, live feed is obtained from an IP camera in the form of frames and the detected faces are passed on to the algorithm. Figure 3.10 shows the HikVision IP camera that we have used to capture livestream. The configuration for the camera has been mentioned previously in section 3.2.1.

Figure 3.10: HIKVISION IP Camera

3.4 Data Collection

In order to train any neural network, we need to have a large dataset. A larger dataset provides more accurate results and overall a better system. To train this system we used IMDB-WIKI [20] dataset which is available for public use on the internet. This dataset consists of 460,723 images. Along with this, we used a local dataset as well.

3.4.1 Setting Up Data for Gender Classification

The dataset includes 460,723 images of celebrities. The faces are labeled using ‘0’ and ‘1’ for male and female faces respectively. During training we used an equal distribution of male and female faces. Figure 3.11 and Figure 3.12 show female and male datasets respectively. Certain images in the dataset are labeled as ‘nan’ and are considered to be neither of the two genders i.e. male or female. Although this dataset consists of many images, during testing it had certain issues when classifying images of local faces. For this reason, we also collected our own dataset of male and female images. Local dataset consists of 33,000 images for both males and females. Dataset was organized by keeping all images of females in a folder named ‘0’, which represents their label. All female images were named ‘0’ and the sequence number of every image in the folder is mentioned in the parenthesis. The name of the folder and images for males was set as ‘1’.

Figure 3.11: Female Dataset

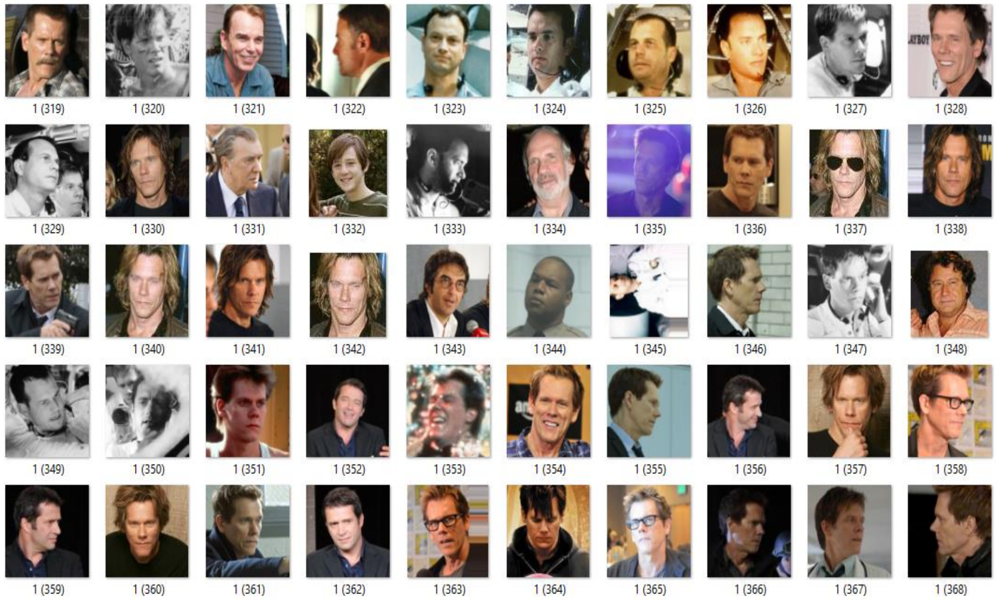

Figure 3.12: Male Dataset

Figure 3.11 and Figure 3.12 show the IMDB-WIKI [20] dataset reorganization for female and male images respectively.

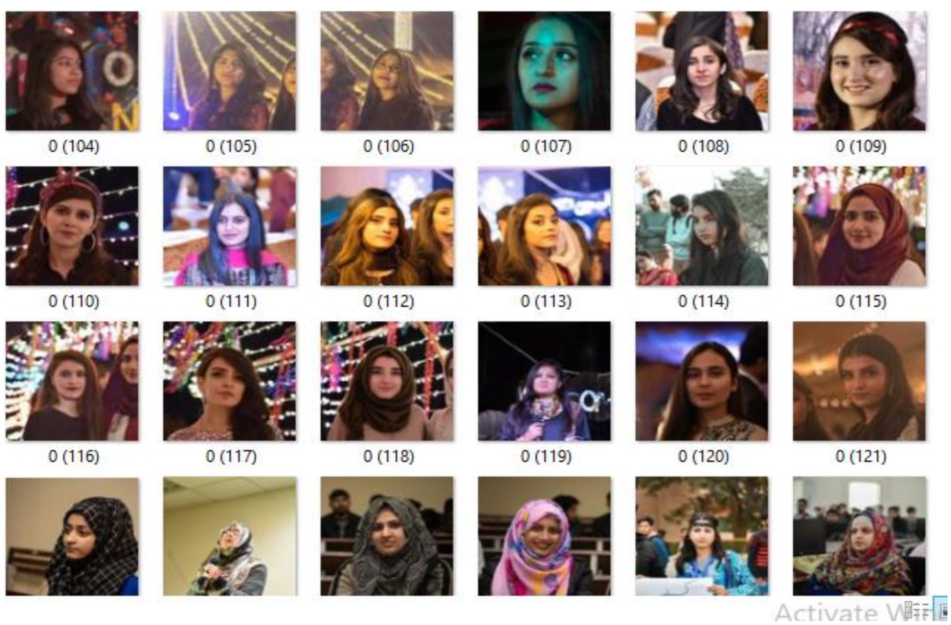

Figure 3.13: Female Local Dataset

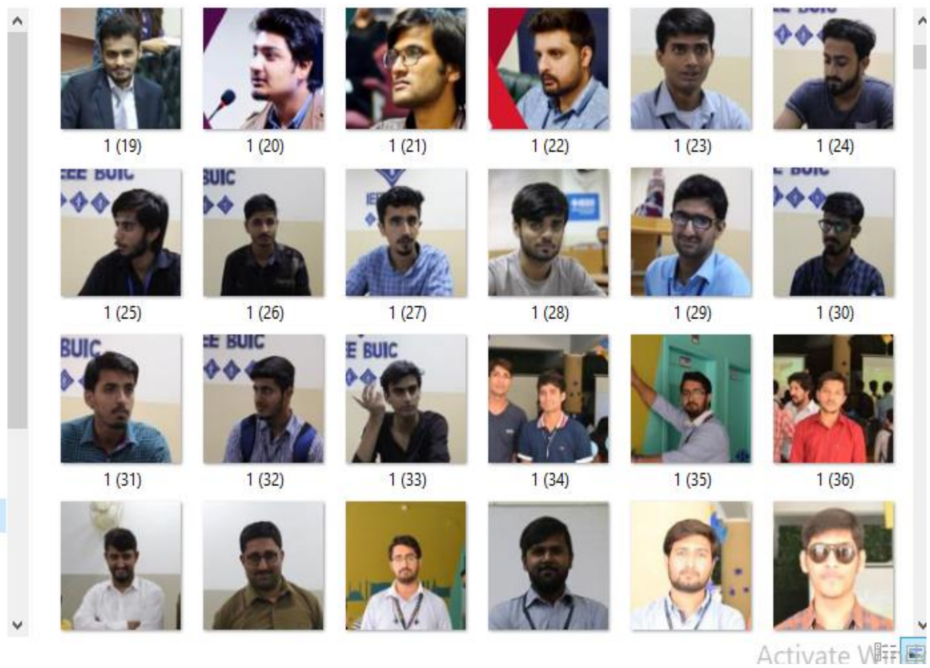

Figure 3.14: Male Local Dataset

Figure 3.13 and Figure 3.14 show the local dataset organization for female and male images respectively.

3.4.2 Setting Up Data for Age Classification

Our system classifies faces into three main age groups:

1. Kids = 0-18 years
2. Young adults = 19-50 years
3. Adults = 50-80 years

The IMDB-WIKI [20] dataset consists of almost equal distribution between the three different age groups. There are 43,200 images of kids, 43,970 images of young adults and 41,330 images of adults. The labels used for the three different age groups are ‘0’, ‘1’ and ‘2’ for kids, elders and adults respectively. We also collected a local dataset for age classification. A total of 33,000 images were used for kids, young adults and elders each with the labels ‘0’, ‘2’ and ‘1’ respectively.

Three different folders were created for each age group and the name of each folder corresponds with the label of the age group. The name of each image is the same as its label and the parenthesis includes the sequence number of the image within its folder. Figures 3.15, Figure 3.16 and Figure 3.17 given below, show the datasets for elders, kids and young adults respectively.

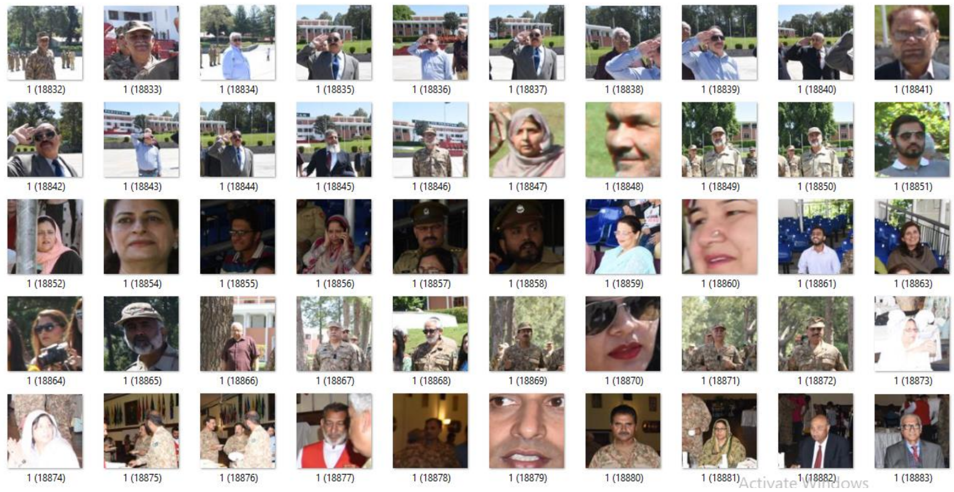

Figure 3.15: Elders Age Group Local Dataset

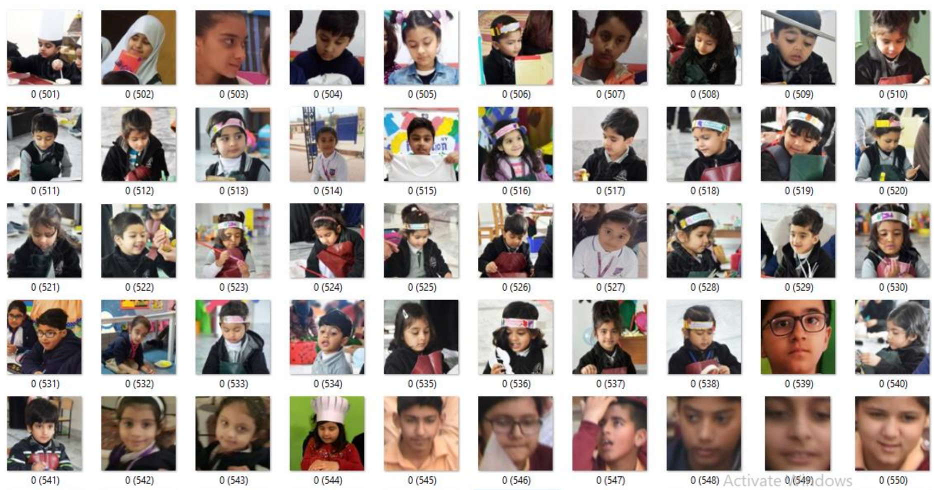

Figure 3.16: Kids Age Group Local Dataset

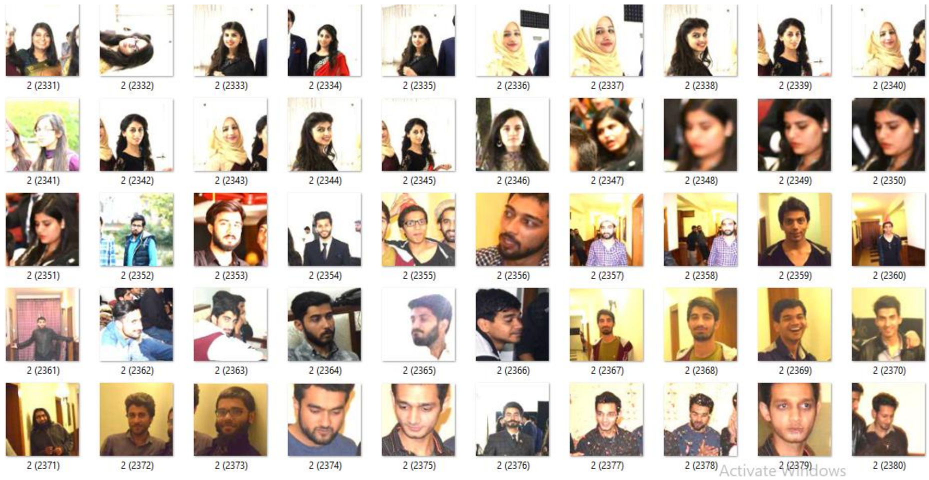

Figure 3.17: Young Adults Age Group Local Dataset

3.4.3 Data Augmentation

Deep learning requires a lot of data. Moreover, we also need regional data. Sometimes however when collecting your own dataset it is difficult to gather a lot of data. For this purpose, we use data augmentation. Described below are some methods of data augmentation:

3.4.3.1 Grayscale Conversion

Grayscale conversion involves converting RGB images with three channels to simply gray images with a single channel. The intensity of grayscale images is stored as an 8-bit integer. This gives 256 different values in the black and white range. The main reason grayscale images are used in image processing is because they require a lot less information to be specified as compared to RGB images. Figure 3.18 below shows simple grayscale conversion results.

Figure 3.18: Grayscale Conversion

3.4.3.2 Contrast Stretching

Image enhancement involves many different techniques. One of these is contrast stretching, also known as normalization. It simply ‘stretches’ intensity values in an image and thus improves the contrast of the image. Contrast stretching requires upper and lower limits. This is the limit over which the images are normalized. Contrast stretching does its best to stretch the intensity values in order to make use of the full range. If the upper and lower limits were defined by ‘a’ and ‘b’ respectively, then contrast stretching can be given by the following formula:

$$P_{out} = (P_{in} - c)(b-a / d-c)+a \tag{3.1}$$

Where ‘c’ and ‘d’ represent the lowest and highest pixel values currently present in the image and ‘P’ is the input pixel.

3.4.3.3 Brightness of Images

This is another simple image processing technique. It simply means making the pixels in an image a little lighter. Brightness can be achieved by the simple addition of a certain number (according to the brightness you desire) to the image pixels. Inversely, the subtraction of a number from the image pixels will result in a darker image.

3.4.3.4 Smoothing of Images

Smoothing is a process done on images to remove noise or add blurring. It creates an image that is less pixelated by passing it through a low pass filter. We applied and compared the effects of two smoothing filters. One being a bilateral filter and the other a gaussian filter. These filters replace a pixel in the image matrix by a weighted average of the surrounding pixels. Results for both filters were somewhat the same, but Gaussian filter proved to be slightly better. The results of Bilateral and Gaussian smoothing filters can be seen in Figure 3.19 and Figure 3.20 respectively.

Figure 3.19: Image Smoothing with Bilateral Filter

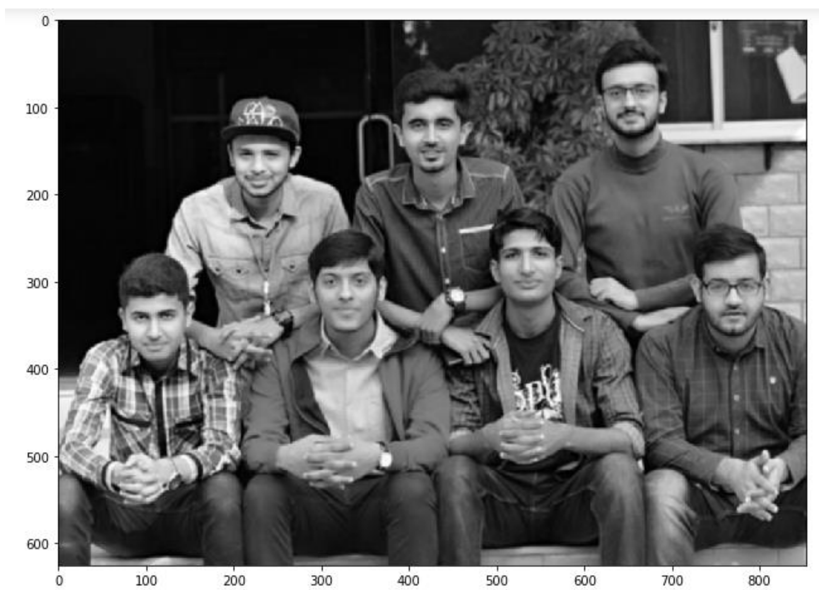

Figure 3.20: Image Smoothing with Gaussian Filter

3.4.3.5 Sharpness of Images

Image sharpening simply helps in highlighting edges and fine details in images. Since the human eye is most often attracted to edges and details in an image it is important to make them as prominent as possible. Also, this helps the system detect faces better

when given a slightly blurry image. Image sharpening is achieved by passing the image through a high pass filter and then replacing this new pixel in the image.
Image sharpening can be represented by:

$$S_{i,j} = x_{i,j} + \lambda F(x_{i,j}) \tag{3.2}$$

Where ‘x’ is the original pixel value with coordinates (i,j) and ‘λ’ is tuning parameter and ‘F’ is the high-pass filter. A higher value of ‘λ’ is used when a more sharpened image is required and vice versa. A gaussian sharpening filter proved best in our case as can be seen from Figure 3.21 and Figure 3.22, when we compare results of Bilateral and Gaussian sharpening filters.

Figure 3.21: Image Sharpening with Bilateral Filter

Figure 3.22: Image Sharpening with Gaussian Filter

3.5 Classification of Gender and Age

After data collection and its pre-processing, the dataset is trained for classification of age and gender groups. Training involves the use of convolutional neural networks. There are a variety of neural networks that are being used currently, to solve different classification problems. Here we have gone with Mini-Xception. This architecture has provided us with a reasonable accuracy.

3.5.1 Implementation of Mini-Xception

Mini-Xception is a combination of the Xception model and "Sequential Fully-CNN" [2] model. The "Sequential Fully-CNN" [2] model reduces the number of parameters by removing the last fully connected layer. The number of parameters is further reduced by removing them from the convolution layers. This step is done through the use of depth-wise separable convolution in Xception model. Depth-wise separable convolutions has two different convolution layers i.e. depth-wise convolutions and point-wise convolutions. The use of depth-wise separable convolution with residual networks helps speed up the computation and alleviates slow processing speed.

3.5.2 Transfer Learning

This method proves to be a big help in deep learning as it allows users to save their time. Instead of training a dataset from scratch, we can make use of pre-trained models. Pre-trained models mean that these models have been previously trained for problems that are similar to the one we are trying to solve. It helps increase performance of a system.

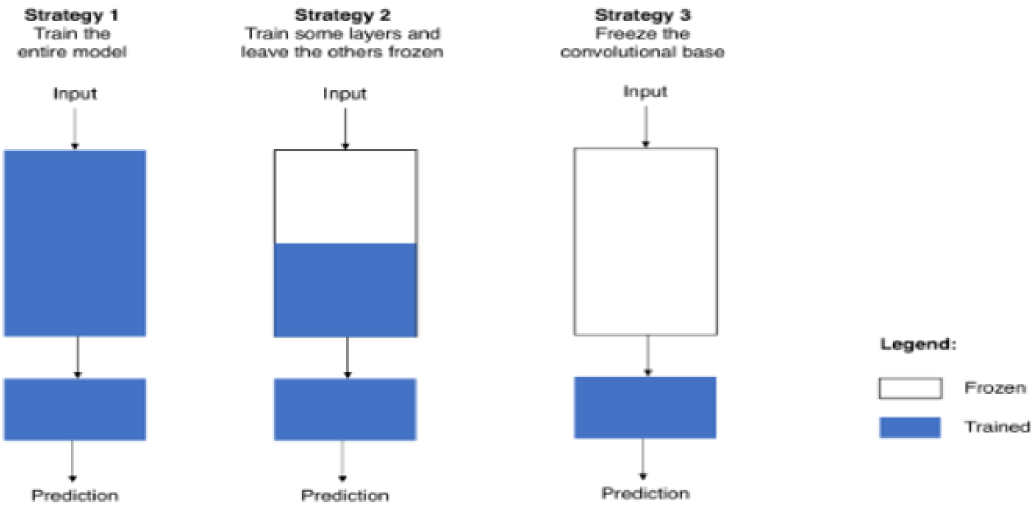

Figure 3.23: Transfer Learning Methods

Figure 3.23 shows that there are three different methods to implement transfer learning. First method involves training the entire model. Second method involves freezing a few layers and training the others and the third method involves freezing all convolutional layers and only training the last layer. As we are using model pre-trained on the IMDB dataset, and our dataset and classification problem our quite similar, we will be using the second and third method.

3.5.3 Training of the Model

The final training was done using Mini-Xception architecture as described above. This architecture results in a reduction of parameters that are 100 times less than a simple fully connected convolutional neural network. This architecture uses global average pooling in the last layer.

3.5.4 Saving the Model and Weights

The weights achieved through training can be saved entirely in hdf5 format. This format makes it easier to handle extremely large amounts of data. Python libraries make it possible for us to save entire models and their weights. These saved models help us test our input images and also live streams from cameras.

3.6 Configuration of libraries

One way to make implementation of code easier is to make use of libraries. Python is a language that provides a large variety of libraries to use. Each library that is installed is equipped with different modules that help you perform many actions without writing any codes. Python libraries that are used for Machine Learning and Deep Learning models are as described below:

3.6.1 Keras

Keras is a Neural Network API. It what helps us build algorithms. Keras installation can be done on anaconda either by directly searching it in site packages and installing it or through the pip install method in the anaconda command prompt as shown in Figure 3.24.

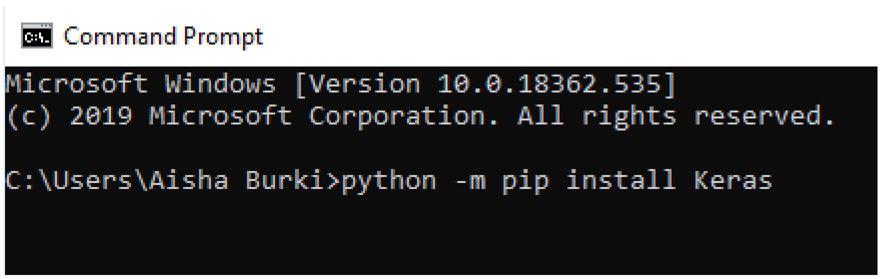

Figure 3.24: Keras Installation

3.6.2 TensorFlow

Tensorflow library helps accelerate machine learning and deep neural networks. It is a graph processor with tensors as its main elements. Installation can be done through command prompt as shown in Figure 3.25 or by searching in site packages.

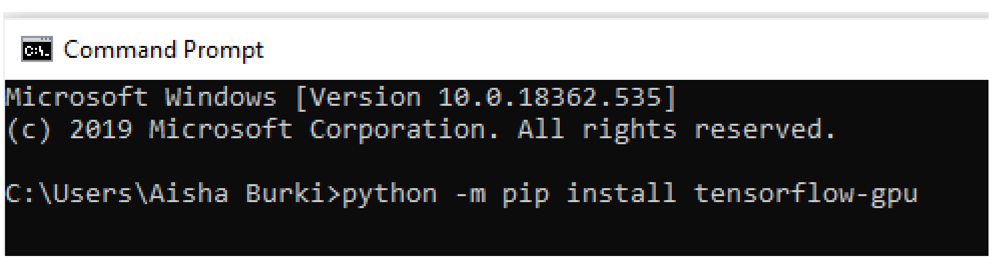

Figure 3.25: Tensorflow Installation

3.6.3 OpenCV

OpenCV is a C++ library of computer vision. It can be installed on anaconda directly or through the pip install method in the command prompt.

3.6.4 Dlib

Dlib is also a computer vision library. It can be installed through similar means as OpenCV as shown in Figure 3.26 given below.

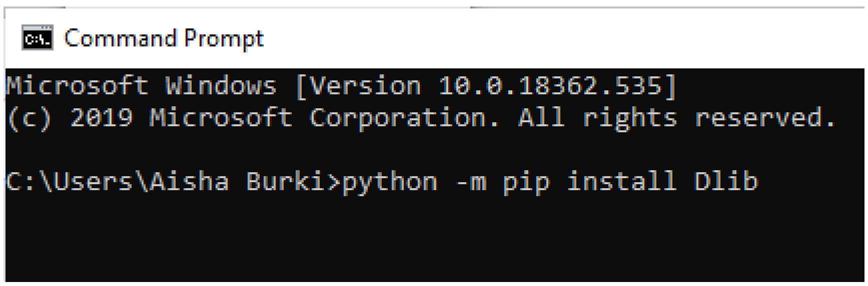

Figure 3.26: Dlib Installation

3.6.5 NumPy

NumPy stands for ‘Numerical Python’. It is also a python package and is used for scientific computing. Its main use is because of its fast computation on arrays and matrices.

3.6.6 Pandas

Pandas is another widely used python library. It provides a 2D table called a DataFrame. This DataFrame has columns and rows. Overall pandas helps creating tables and plotting graphs and calculating columns based on other columns.

3.6.7 SciPy

SciPy is a python library that can be used for optimization, linear algebra and signal and image processing. Its method of installation is similar to other python libraries.

3.6.8 Scikit-learn

This is a machine learning library for python. It has many classification, regression and clustering algorithms.

3.6.9 PIL

The python imaging library (PIL) is one of the most important libraries. It helps in opening, saving and making changes in different images.

3.7 Web Application Introduction

After the completion of our model, the next step was to provide a user friendly interface to our clients. For that purpose we developed our web application which we have named “C-Counter”. Figure 3.27 shows the home page of our web application that will be viewed by customers when they first open the website. After logging in, it will help our clients to view the statistics of customers that have walked in to their stores.

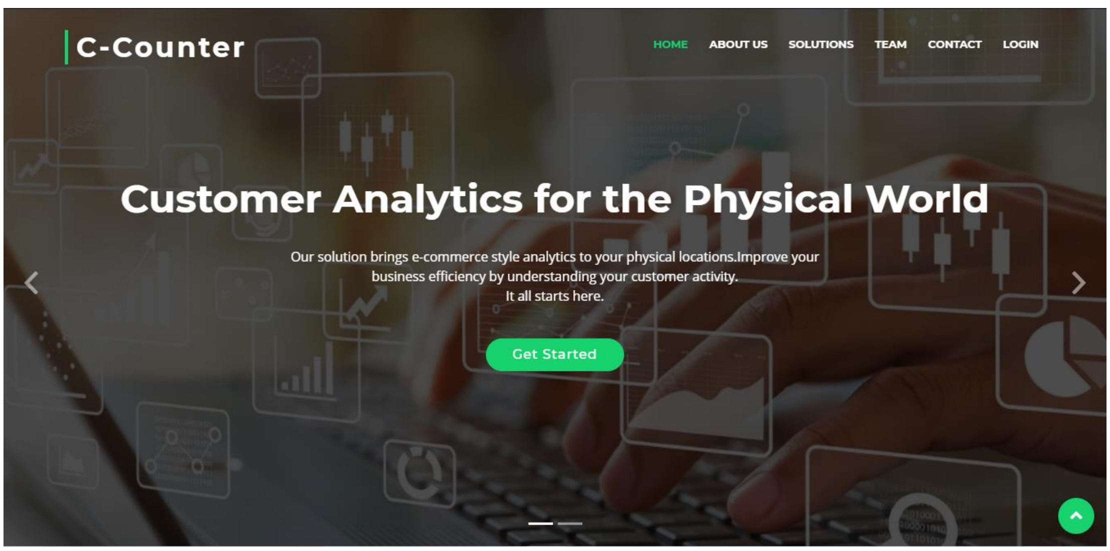

Figure 3.27 C-Counter

All the data of customers will be managed by our Database Management System. The most commonly used database management systems are Oracle data, SMS and Microsoft access. We have used MySQL in creating database for our system. It’s a domain specific language used in managing database where relational data is being used. MySQL queries are used in updating and retrieving data from the rows and columns of database.

Figure 3.28 Microsoft SQL Server

3.7.1 Connection with MySQL

SQL connection request is made to enable other codes to operate with database. In our client/server database system this connection is somewhat similar to the network connection to server. SQL credentials are used to initiate SQL connection object which connects it to database.

The credentials that we have deployed in SQL server are:

i. Host name
ii. Password
iii. Database name
iv. User name

3.7.2 MySQL Queries

Queries are used for updating and obtaining data from database. It enables user to select particular data from database. Whenever we require data from database, we send request for retrieving that data in the form of query into our database management system. The queries that we used in our server are given below.

Create Database:

Create Database query was used to create our database in SQL server which we have named as ‘**fyp**’.

Create Table:

Create Table query was used to create four tables in our database. Figure 3.29 shows the table that maintains gender ratio, Figure 3.30 shows table that maintains age group ratio, while the total number of people and customer account information is shown in Figure 3.31 and Figure 3.32 respectively.

←T→	id	gender	number
Edit Copy Delete	1	Male	0
Edit Copy Delete	2	Female	3

Figure 3.29 Table for Gender Ratio

←T→	id	age	number
Edit Copy Delete	1	Kids	0
Edit Copy Delete	2	Elders	3
Edit Copy Delete	3	Young Adult	0

Figure 3.30 Table for Age Group Ratio

←T→	id	number
Edit Copy Delete	1	3

Figure 3.31 Table for Total People

←T→	id	username	password	email
Edit Copy Delete	1	admin	admin	admin@test.com

Figure 3.32 Table for User Information

Below is the list of tables in database

i. count_gender

ii. count_age_group

iii. count_total

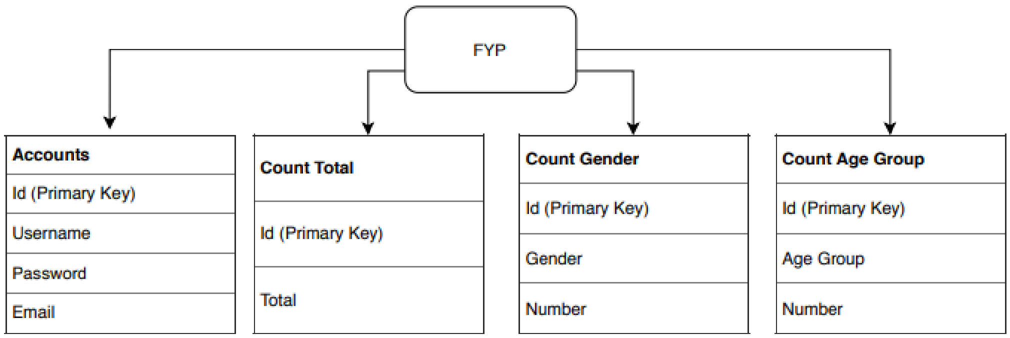

Figure 3.33 Database Structure

Figure 3.33 provides an overall view of all four table included in the database, that have also been mentioned previously.

3.8 XAMPP Server

XAMPP is a free and open-source cross-platform server. Its package comes with Apache Web server, SQL database and PHP. It is very handy in creating database and PHP code in windows platform. Its programing languages include PHP and pearl. Connection between PHP and HTML is made easily without any further configuration. XAMPP Server operates in Windows as well as in Linux.

3.8.1 Control Panel

We can view and change the settings of XAMPP by using XAMPP control panel, shown below in Figure 3.34. It can also configure and manage all the available XAMPP services.

Figure 3.34 XAMPP Control Panel

XAMPP control panel consist of three main sections

i. Modules
ii. PID(s)
iii. Port(s)

Module section is for the services that are available in XAMPP. The services that Module section includes are:

i. Apache
ii. MySQL
iii. FileZila
iv. Mercury
v. Tomcat

These services offered by XAMPP are accessed by clicking the start button and they are configured by clicking on config button that is located in control panel as can be seen in the Figure 3.35 below.

Figure 3.35 Access to Modules of Control Panel

Port(s) number shows the port number that is dedicated to the intended service of XAMPP control panel. For example in Figure 3.35 above, Apache is using port number 80,433 and MySQL is using 3306. PID(s) are the port IDs designated to each service.

3.8.2 XAMPP Server Configuration

We followed the steps mentioned below for configuring XAMPP server in PC:

I. We downloaded and installed the XAMPP server then configured its settings.

II. From start menu of Windows, we run XAMPP server. This launched the control panel of XAMPP sever from where we accessed the services that we needed.

III. We clicked on start button of Apache service to deploy and manage the overall functioning of the web server.

IV. Similarly for acquiring database services we clicked on start MySQL button.

3.8.3 Connection of Front-End with Database

This connection was made simply by using php, which is a scripting language for website development. We connected our website’s dashboard to database using the following queries:

```
$query = "SELECT gender, number FROM count_gender";
$query = "SELECT age, number FROM count_age_group";
$query = "SELECT number FROM count_total";
```

These queries were used to retrieve the data and show it to on our web application for the client. Gender, age and number are the names of columns in the tables while, count_gender and count_age_group are the tables where these columns are present. Similarly we also used php to connect our login page to the database. Using database, we ensure that the username and password entered by the user were correct and after verification, we then display the dashboard for the user to see.

3.9 User Interface

As described earlier the purpose of creating web services was to provide users with a friendly interface where they can access and view important information. We have created dashboard for users to access and view data and also to navigate the results.

3.9.1 Tools and Techniques

There are many tool and techniques present for creating web pages and dashboards. The tools and techniques that we used in our project for creation of dashboard are as under:

- Bootstrap
- jQuery
- Java script

Bootstrap

Bootstrap is a framework that is used for front end developing and designing of web application. It can integrate with HTML, CSS and Java script code. The templates of HTML and CSS are available for developers in this framework.

Bootstrap CDN

In our project we included bootstrap from design section’s content delivery network. MaxCDN provides CDN support for Bootstrap CSS and JavaScript.

Instead of downloading and hosting CDN we followed the steps mentioned below to insert bootstrap in our HTML page:

I. We copied bootstrap file link from bootstrap official site present at web.
II. We pasted the copied link in the head section of HTML file.

jQuery CDN

Irrespective of bootstrap we downloaded and hosted jQuery CDN. From CDN (Content Delivery Network) we added it in our dashboard. It requires all libraries related to it be installed in your system. So we included all libraries through HTTPS protocol.

We followed the steps mentioned below to include jQuery in our HTML page:

I. We copied jQuery file link from bootstrap official site present at web.
II. We pasted the copied link in the head section of HTML file.

Java script

Java script is a tool used for enhancing the performance of web pages. It is a light weight dynamic programming language used as a part of web pages.

We followed these steps to include js files in our HTML page:

I. We copied js file link from official site present at web.
II. We pasted the copied link in the head section of HTML file.

These were all advanced level tools and languages that we used in our project for designing web application and dashboard of our system.

3.9.2 Web Page

Our web page is designed for our clients to access the information, to get support from us and also for new users who can view our project services and to get information they need to excel in their business. The menu of our web page consists of five main sections.

3.9.2.1 About Us

This section takes users to a page where we have stated our mission, plan and vision. The About Us page can be seen in Figure 3.36. It gives an idea to new users to help them get information about our product and why they require it to boost their business.

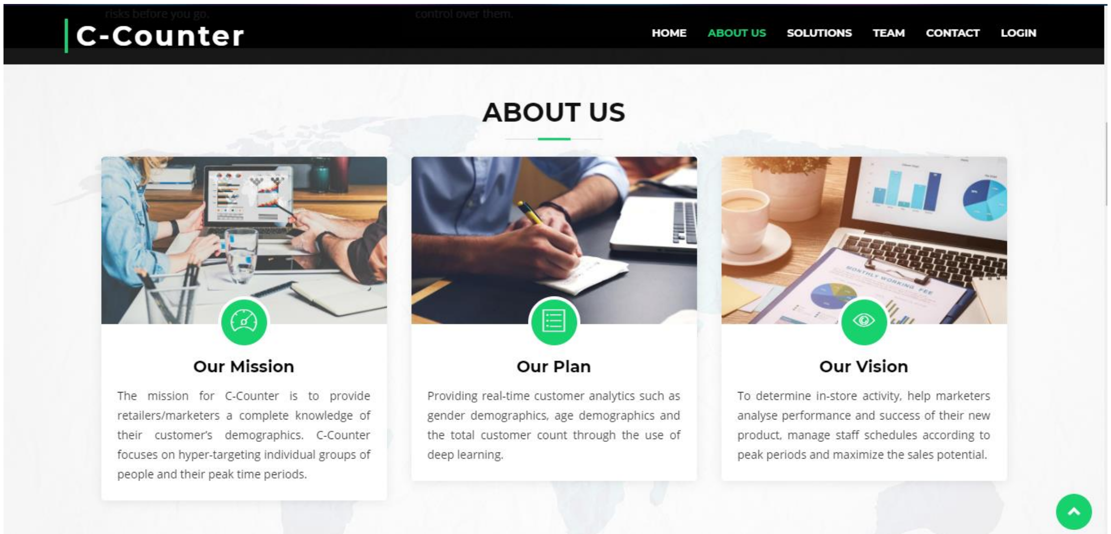

Figure 3.36 About Us Page

3.9.2.2 Solutions

All the solutions that we are offering through our product are described in this section, shown in Figure 3.37. It helps new users get familiar with our product solutions and usage.

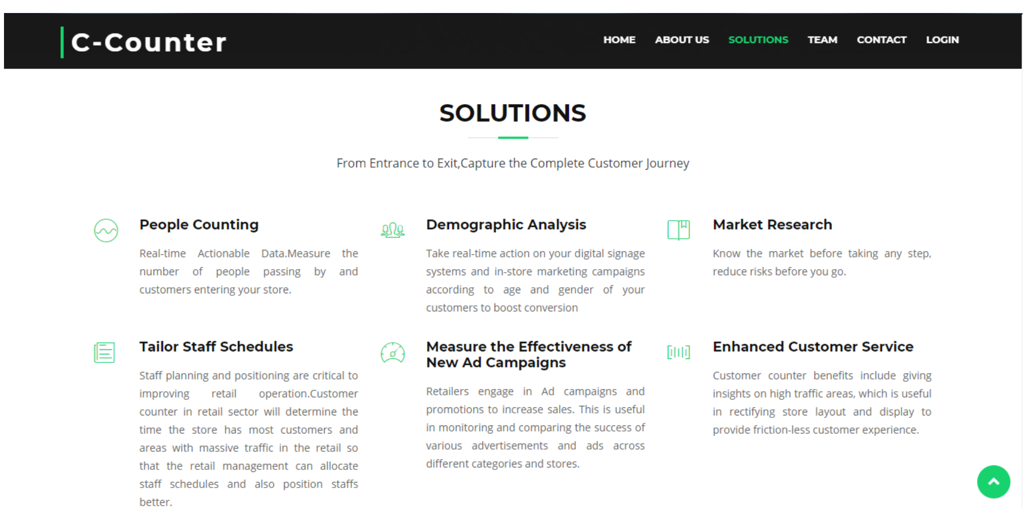

Figure 3.37 Solutions Page

3.9.2.3 Team

This section is solely dedicated for the audience to get information about our project team. Clients and new users can directly reach and check the social accounts of our project team. The Team page can be seen in Figure 3.38 given below.

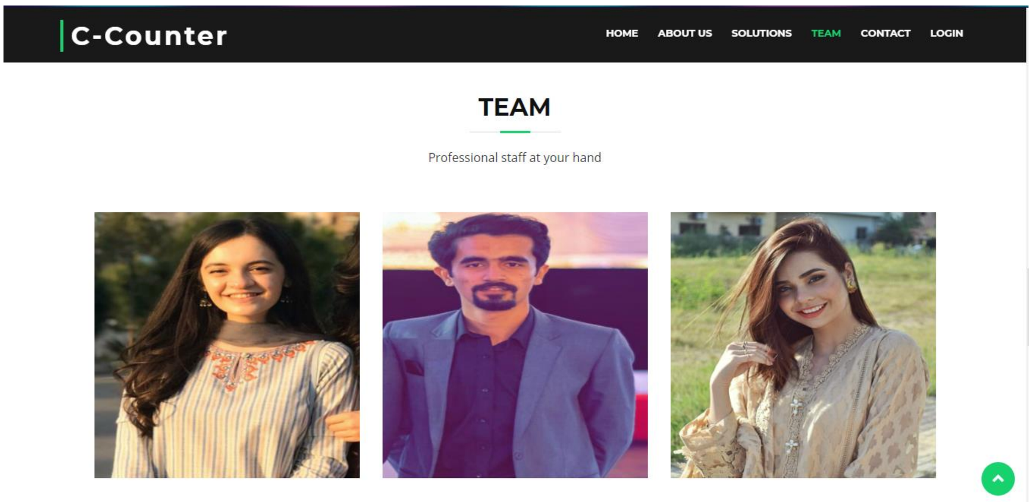

Figure 3.38 Team Page

3.9.2.4 Contact Us

This section, shown in Figure 3.39, is for those people who are interested in working with us and want to install our system in their outlet and stores. Complete information about how to contact and reach us is provided in it.

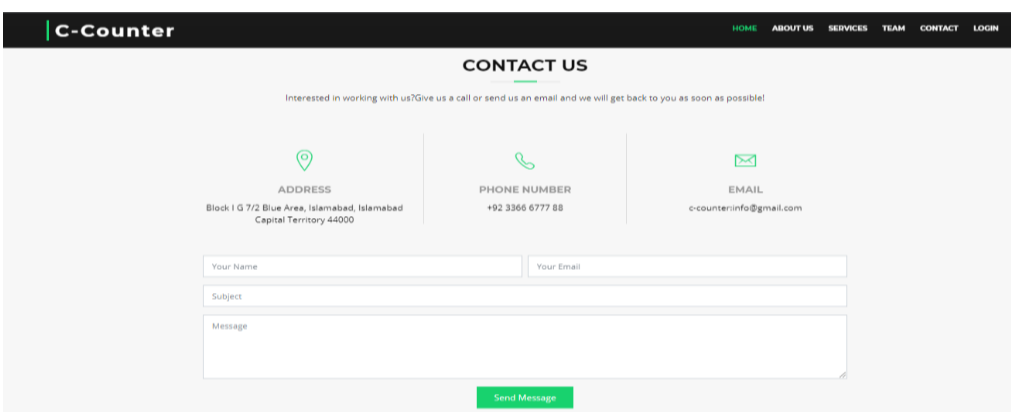

Figure 3.39 Contact Us Page

3.9.2.5 Login

This option is for our clients who have installed our system in their outlet and we have provided them a login ID and password, which they can enter on the login page as shown in Figure 3.40. Through this they can access the dashboard and analyze the information of customers who have visited their store.

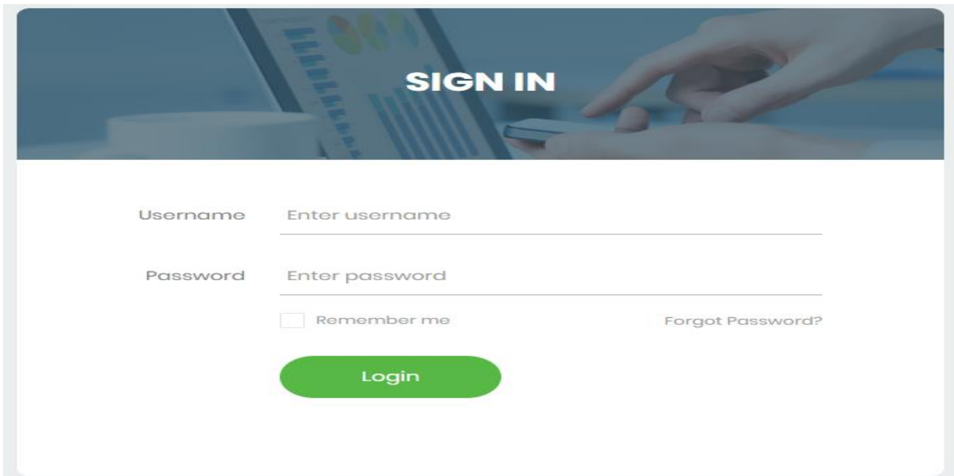

Figure 3.40 Login Page

3.9.3 Components of Dashboard

In this section we will discuss the modules of our user login dashboard. The dashboard is the page that the user sees immediately after logging in. this is shown in Figure 3.41. We have created a simple dashboard that lets the user know the age and gender demographics present in the store currently.

The demographics for both gender and age groups are displayed using pie charts. The dashboard refreshes after every few seconds, so that the customers that are being counted in real time, keep showing up on the dashboard. Besides this, the dashboard also includes an option for viewing the livestream from the IP camera. This option can be seen in Figure 3.41 and upon clicking this option another page opens on which the livestream can be viewed, as shown in Figure 3.42.

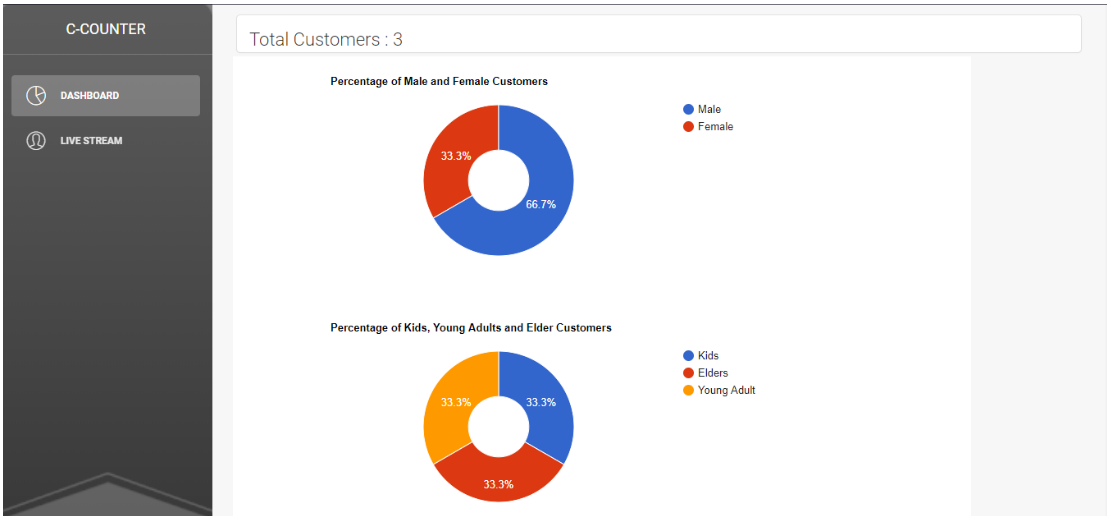

Figure 3.41 Dashboard

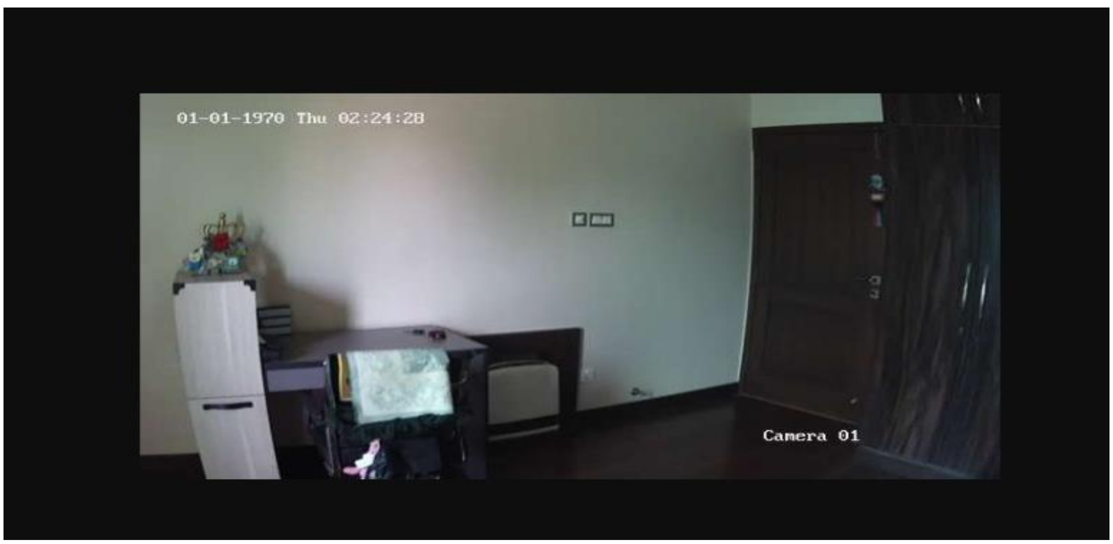

Figure 3.42 Live Stream

Note: Due to Covid-19, IP Camera could not be tested in a local store.

3.10 Google Visualization APIs

It is a Java script API that is used for accessing and viewing structured data using Java Script in web pages.

Advantages

- It is very powerful and compatible.

- It is open source chart library.
- It can display live data in form of interesting charts.
- It has a vast amount of chart gallery.

Procedure to Use Google Charts

I. We copied the reference of JavaScript file link from official site.
II. We pasted the copied link in the head section of HTML file.
III. We loaded the Google Chart Libraries in head section.
IV. We used draw() function to display chart in HTML section.
V. We instantiated the chart object and after that we drew chart by using draw() function.
VI. In the argument of draw() function we passed our data that is to be displayed.

3.11 Connection with Python Testing

Connection of python code with our database was fairly simple as well. As we are using MySQL database, we need to make a connection to this database in python. This was done by installing the MySQL and MySQL Connector Python packages through pip command. After installing all necessary packages and libraries, data can be saved from python code into database using simple MySQL queries, such as INSERT and UPDATE. It is necessary to use commands such as connect( ), before entering data, execute( ) command for entering data and close( ) command after data has been sent into database. We made use of three different functions to enter the data of gender, age group and total people. Each function accessed its respective table in the database and made updates accordingly.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Real Time Testing

We began our experimentation phase by testing our entire system in real time through our IP camera. For that purpose we extracted the region of interest, in our case this is faces from the frames coming from the footage of our IP camera and then applied age and gender classification algorithms on them. The extracted results are shown in the next section.

4.1.1 Face Detection

Extraction of region of interest i.e. real time face detection was the first step that was involved in our real time testing. The footage was taken from our IP camera, from which the frames were forwarded to our Dlib face detector. Each face once detected by our face detector is then moved towards the classifier for rest of the procedure. After classification a bounded box is made around the detected faces using width, length and height of the faces which can be viewed in live footage.

4.1.2 Gender Classification

After face detection through Dlib, the detected faces are now transferred into correlation factor. The purpose of this is to only pass unique faces into the classifier as there must be no repetition of the faces. The unique faces are then passed to classifier for prediction. After prediction, labels are assigned to the detected faces and on the basis of those assigned labels, the predicated class is being displayed on the output video i.e. male or female on the live footage.

Meanwhile a counter is being incremented which is counting the strength of males and females and updating the database accordingly in real time. Testing of gender classification on a static image has been shown in Figure 4.1. Testing in real time through video footage has been shown in Figure 4.2 and Figure 4.3 given below.

Figure 4.1 Testing w.r.t Gender Example 1

Figure 4.2 Real Time Testing w.r.t Gender Example 2

Figure 4.3 Real Time Testing w.r.t Gender Example 3

4.1.3 Age Group Classification

Similarly by using IP camera we extracted faces and by using correlation function we obtained unique faces and applied age group classification. The unique faces in this case are passed into age group classifier. We formed bounding box on the detected faces in live footage and on the basis of the predicted class, displayed labels on them i.e. kids, young adults or elders in live feed.

Just like in previous case a counter here is also maintaining the figure of how many kids, young adults and elders faces are being detected and also the total people detected. Testing of age group classification on a static image has been shown in Figure 4.4. Testing in real time through video footage has been shown in Figure 4.5, Figure 4.6 and Figure 4.7 given below. Instances involving eyewear and faces covered by masks are also shown in these figures.

Figure 4.4 Age Group Testing

Figure 4.5 Real Time Testing w.r.t Age Group Example 1

Figure 4.6 Real Time Testing w.r.t Age Group Example 2

Figure 4.7 Real Time Testing w.r.t Age Group With Mask

4.2 Results

We performed testing on images for gender and age group classification in order to calculate the testing accuracy of our classifier. The accuracy received through Gender classification was 80% and for Age Group Classification 83%.

4.2.1 Confusion Matrix

We are using Confusion Matrix in order to describe the performance of our system on the data set. The results of both classifiers that we used are as under:

Gender classification

To get results, we used 8,024 images. All falsely and correctly predicted images can be seen below:

Table 4.1: Gender Classification Confusion Matrix

	Male	**Female**
Male	**True Positive** 2568	**False Positive** 821
Female	**False Negative** 597	**True Negative** 2430

Age Group classification

Table 4.2: Age Group Classification Confusion Matrix

Predicted Labels

Actual Labels

	Kids	**Elders**	**Young Adults**
Kids	383	0	107
Elders	19	179	300
Young Adults	15	7	495

4.3 System Validity

Frame rate per second (FPS):

In an IP camera, frame rate per second refers to number of frames being recorded by IP every second. This term is being used in films, games, videos etc. In our case it is very important to have the knowledge of fps of the live footage that is being displayed at our system screen as well of that of the testing frames.

The formula that we used to acquire the fps of the live camera feed is:

FPS= (start time - current time)/Total number of frames

As seen from the formula above, we are calculating it by subtracting the start time from the current time and dividing it by the total number of frames. Our frame rate per second with IP Camera video testing was of **25 frames per second**.

CHAPTER 5

CONCLUSION

On the basis of the work that we have done we can conclude that the accuracy of available data and the environment from which we are getting the live feed are the determining factors of reliability and accuracy of our model. Our model will give more accurate results if the quality and resolution of live feed are enhanced. Moreover, our model enables our client to view and analyze the complete journey of their customers from the moment they walk into their stores to when they exit. It will enable them to know the demographics of their customers so they can devise new marketing strategies to flourish their business.

5.1 Recommendations

Our research has revealed that accuracy of model depends upon the quality of data collected and camera being used. So in order to provide clients with good and accurate results camera with high resolution must be used. In this way overall performance of system will improve. Similarly while making the local dataset good quality camera should be used, this will help in extracting more features and regions of interest. In this way training of system will be enhanced. As we all know there is always a room for improvement, the overall system performance can be boosted by acquiring the latest and advanced machine learning tools.

5.2 System Limitations

The reliability and precision of results depends upon the quality of live feed. System performance will be affected to those areas where there is no proper lighting arrangement for capturing the live feed. Under these circumstances system will not be able to show accurate results. In gender classification, system will be unable to identify the gender of women who are wearing veils as the system is not familiar to this type of dataset.

CHAPTER 6

FUTURE WORK

The basic idea was to devise a system which will enable the owner of superstores to have complete knowledge about their customers' age and gender group. A lot of improvements can be made on this project. Some of them are listed below.

1. System must be trained on a dataset which will enable it to detect the gender of women wearing veils.

2. A feature must be introduced that will enable the outlet owner to view how often a customer visits his store and purchases something. Thus providing discount deals and vouchers to the customer based upon how often they visit.

3. This system could be installed in a shopping mall to monitor the age and gender group on different floors of mall. The advertisement on digital screen of malls could be done accordingly to the majority age and gender group on that floor.

4. The system could be deployed in Marquees and Wedding Halls to keep the track of number of guests to avoid any inconvenience.

5. The system accuracy can be further improved by training it on a bigger dataset than the one which we have used.

6. The system must be tested under different lighting effects and environment conditions.

REFERENCES

[1] Sumit Saha. 2018. A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networksthe-eli5-way-3bd2b1164a53 (Accessed on 5/08/2019)

[2] Arriaga, O., Valdenegro-Toro, M., & Plöger, P. (2017). Real-time convolutional neural

networks for emotion and gender classification. arXiv preprint arXiv:1710.07557.

[3] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition.

[4] Vikas Gupta. 2018. Face Detection – OpenCV, Dlib and Deep Learning (C++ / Python). https://www.learnopencv.com/face-detection-opencv-dlib-and-deep-learning-c-python/ (Accessed on 15/08/2019)

[5] Sharma, Manik., Anuradha, J., Manne, H. K. and Kashyap, G. S. C. (2017). Facial detection using deep learning. Published at IOP Conference Series: Materials Science and Engineering, Volume 263, Issue 4, pp. 042092 (2017).

[6] Arun Ponnusamy. 2018. CNN based face detector from dlib. https://towardsdatascience.com/cnn-based-face-detector-from-dlib-c3696195e01c (Accessed on 2/09/2019)

[7] Kholod,T., and Balkan,S. (2014) Is Video Analytics a Game Changer for Market Research?

[8] Zhang, K., Sun,M., Han, T X., Yuan, X., Guo, L., and Liu, T. (2016). Residual Networks of Residual Networks: Multilevel Residual Networks

[9]Antipov, G., Baccouche, M., Berrani, S A. and Dugelay, J L. (2016). Apparent Age Estimation from Face Images Combining General and Children-Specialized Deep Learning Models.

[10] Zhang, Zhiqin. (2016). Apparent Age Estimation with CNN. 2016 4th International Conference on Machinery, Materials and Information Technology Applications

[11] Rothe, R., Timofte, R., and Gool, L V. (2016). Deep expectation of real and apparent age from a single image without facial landmarks.

[12] Smith. P., and Chen, C., (2018). Transfer Learning with Deep CNNs for Gender Recognition and Age Estimation

[13] Deghan, A., Ortiz, E G., Shu, Guang. and Masoon, S Z. (2017) DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network

[14] Liu, X., Li, J., Hu, C., and Pan, J., (2017). Deep convolutional neural networks-based age and gender classification with facial images, 2017 First International

Conference on Electronics Instrumentation & Information Systems (EIIS), Harbin, 2017, pp. 1-4.

[15] Hebda, B. and Kryjak, T. (2016). A compact deep convolutional neural network architecture for video based age and gender estimation.

[16] Sindagi. V A. and Patel, V M. (2017). CNN-based Cascaded Multi-task Learning of High-level Prior and Density Estimation for Crowd Counting.

[17] Nogueira, V., Oliviera, H., Silva, J A., Vieira, T. and Oliveira, K. (2019). RetailNet: A Deep Learning Approach for People Counting and Hot Spots Detection in Retail Stores, 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)

[18] Revathi, T. and Rajalaxami, T.M. (2019) Deep Learning for People Counting Model.

[19] King, D.E. (2015). Max-Margin Object Detection Davis E. King.

[20] Rasmus Rothe, Radu Timofte and Luc Van Gool IMDB-WIKI – 500k+ face images with age and gender labels https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/ (Accessed on 12/07/19).

ABBREVIATIONS

CNN: Convolutional Neural Network

DenseNet: Densely Connected Convolutional Neural Network

AI: Artificial Intelligence

OpenCV: Open source Computer Vision

ReLU: Rectified Linear Unit

RGB: Red Green Blue

HOG: Histogram Of Gradient

SVM: Support Vector Machine

MMOD: Maximum Margin Object Detector