CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF THE STUDY

Human's face reveals various information including gender, age and ethnicity. They provide important cues for many applications such as biometric authentication and intelligent human-computer interface. Recently many applications from biometrics, security control to entertainment use the information extracted from facial images that contains information about age, gender, ethnic background and emotional state. Automatic age and gender estimation from face images is one of the popular and challenging task that have different and diverse fields of applications.

Age and gender recognition has long been recognized as an important module for many computer vision applications such as human-robot interaction, visual surveillance and passive demographic data collections. Determining the age and gender of individuals from a live camera has many applications in the world of advertising. When a frame is captured, a face detection technique is employed to detect the face object. After detecting and aligning the face, the information allowing for greater facial discrimination is extracted and these data are then used to recognize the face's age and gender category

Interestingly, research has shown that age estimation and classification are affected by gender differences as well as actual age. Indeed, both facial and gender classification have been studied together as a related problem. Similarly, the two problem have been tackled simultaneously in other fields such as automatic speech recognition. Like other branches of facial analysis, automatic aging and gender classification are hindered by a host of factors including illumination variation, facial expressions, and pose variation to mention but a few. Several approaches have been documented in the literature to circumvent these problems. Facial aging can be categorized into age estimation, age progression, and age invariant face recognition (AIFR). Age estimation refers to the automatic labeling of groups or the specific ages of the individuals using information or features obtained from their faces. Age progression reconstructs the facial appearance with natural aging effects, and AIFR focuses on the ability to identify or verify people's faces automatically, despite the effects of aging. Gender classification automatically assigns on of the two sex label (male/female) to a facial image. Studies have shown that we humans are able to differentiate between adult male and female faces with up to 95% accuracy. However, the accuracy rate reduces to just above chance when considering child faces.

1.2 STATEMENT OF THE PROBLEM

The task is to predicts the age and gender of a person from his facial attributes using Adience dataset which can be done with the following four components;

- The first component is facial detection; which detects the face and minimize the lighting effect.
- In the second component features (attributes) are extracted and unwanted features are removed.
- In the third component; the required features are selected.
- And the final component; (which produce the result) is a classification where the age and gender will be estimated.

Thus, two separate problems will be studied: face recognition and determining the corresponding age and gender of face objects.

1.3 AIM AND OBJECTIVES

Aim:

In this project, we focused on face detection and recognition capabilities that will identify age and gender of a person in an image or by opening laptop webcam (in real time).

Objectives:

✓ To establish an image recognition system that can estimate the age and gender from a frontal facial images or real-time video captured from webcam.

✓ To create a python application using Open Source Computer Vision Library for python (OpenCV-python) and deep learning module to perform real-time age and gender detection.

- ✓ To detect the face in a given frame, extracts features related to the face objects and uses deep learning model to produce the result.
- ✓ To propose a methodology for the automatic detection of face and recognition of facial/pattern occurrences from sample images captured in real time.

1.4 SCOPE AND LIMITATION

Scope:

This project focuses in the problem of age and gender classification of an image, and the experiment is conducted on the Adience dataset, which is designed for age/gender classification in an unconstrained environment. The Adience dataset contains 19,487images of 2,284 subjects with 8 age groups: 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53 and 60-. Most age groups have around one to two thousand images except for two senior groups (only around eight hundred images each), and the 25-32 group (about five thousand images).

Limitations:

There will be some restrictions which includes:

- The project was not predicting with high accuracy because of limited number of training and testing data.
- Using of Classification instead of Regression because of no enough training data.
- The predicted age is grouped instead of predicting discrete value.
- The proposed system is not a fully automated system.
- The face detection is based on the frontal face only.
- As a standalone application, it does not require installation.

1.5 SIGNIFICANCE OF THE STUDY

This technology has a broad scope and the potential to make a large impact, which could be used to aid assisted vision devices for those with deteriorating or lose eye sight. Social media like Facebook could use information about age and gender of the people to better infer the context of the image and it could also be used in Human computer interaction.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter talks about the review of related works on age and gender estimation method. Also it makes a discussion on all the reviewed literatures.

2.2 REVIEW OF RELATED WORKS

The application of machine learning has rapidly advanced in various fields such as speech recognition, text and image recognition and so on.

An example of image classification using machine learning is demonstrated by Pratik Devikar (2016). He used the pretrained model, Inception v3 to do transfer learning. The dataset used to retrain the model is obtained from Google images. The aim of this research was to train a model which can recognize and differentiate 11 types of dog breeds. Hence, he prepared 11 types of datasets, each datasets comprised of 25 slightly different images of particular dog breeds. To ensure the uniformity of the datasets, the images were set to a resolution of 100×100 pixels. Throughout the experiment, he implemented Python programming language and import TensorFlow library to conduct classification task. The accuracy score was generated by using SoftMax algorithm. The resulting accuracy of testing he achieved reached 96% (Devikar, 2016).

Another example that utilised machine learning in image recognition is demonstrated by Tapas (2016). The aim of this experiment is to classify plant phenotyping. He used the pretrained model, GoogleNet to do retraining. The dataset was extracted from the database of Computer Vision Problems in Plant Phenotyping (CVPPP 2014) database. The dataset comprised of 3 categories, 2 on Arabidopsis, with 161 images and 40 images respectively. Another category of the dataset was Tobacco species, which consisted of 83 images. The retraining process is conducted via TensorFlow library and python as a programming language. The output was displayed in probability using the computation of SoftMax function. The results of accuracy based on testing image reached 98%.

In addition, a similar study on the flower classification using Inception v3 worth a consideration. The study was based on Inception v3 model of TensorFlow platform. The experimental datasets were acquired from two sources, Oxford-17 database, which consist of 17 categories of flowers and Oxford-102, which consist of 102 categories of flowers. The results depicted by SoftMax function regarding the possible output with the input of testing mages are compared according to the two types of dataset. The result shows model trained under Oxford-17 dataset reach 95% of accuracy whereas Oxford- 102 dataset gives an accuracy of 94% (Xia & Nan, 2017).

According to Chin et al. (2017), a research on intelligent image recognition system for marine fouling using SoftMax transfer learning and the deep convolutional neural network was done. They implemented transfer learning by retraining Google's Inception v3 model and SoftMax as an output of prediction based on image input. The images were processed by Open Source Computer Vision Library (OpenCV) and the retraining process is done with the help of TensorFlow Library. At the beginning of the process, Raspberry Pi 3 captured image of the marine fouling. The image was then uploaded to cloud to be classified by the retrained Inception V3 model and convolutional neural network. Then, the image was processed and the percentage of the area of macro fouling organisms was determined. The percentage in the range of 25-40% was considered as heavy fouling and cleaning process must be conducted. The datasets were obtained from captured images from the web. The model was retrained to classify 10 classes of fouling species, with dataset size in the range of 82-228 images. In order to enhance the accuracy of the model, the model was trained twice. Results show the lowest improvement in percentage is 10.302% where the highest can reach 41.398% of improvement. Upon testing on the reliability of the trained model, the highest accuracy achieved among the 10 classes of fouling species was rock oysters, which can reach 99.703% correct prediction. On the other hand, finger sponge species possessed the lowest accuracy, which is 76.617%.

According to Tamkin et al (2013), they claimed that diabetic retinopathy can be detected with the application of deep Convolutional Neural. They extracted a dataset from Kaggle competition database. The database was chosen because the images are taken in various conditions, including different cameras, colours, lighting and orientation. The more variety was the images sources, the higher the robustness of the trained model. A total of 35,126 images, with the size of more than 38 gigabytes is separated in the ratio 0f 8:2, whereby 80% of the images are used as training set, 20% are used as testing set. All images were resized to 256 pixels x 256 pixels. The highest accuracy achieved at the end of the experiment is 92.59%.

According to Thukral, et al (2012), they proposed a hierarchical method to estimate human age. The datasets were obtained from FG-Net website. Upon gathering the dataset, they grouped the images into 3 major groups, in the ranges of 0-15, 15-30 and 30+years old respectively. The experiment can be divided into 3 steps, feature extraction, regression and classification. In feature extraction, facial landmarks points at corners or extremities of eyes, mouth and nose are extracted. Regression was conducted by determining the independent variable, x and the dependent variable, y. Next, they used the Relevance Vector Machine (RVM) regression model conduct machine learning according to the age groups. After that, in classification phase, they utilised 5 types of classifiers, including μ-SVC, Partial Least Squares (PLS), Fisher Linear Discriminant, Nearest Neighbor, and Naïve Bayes to classify the images into the correct age group. Results showed that if the classifiers were able to classify the images into correct age group, the age estimation task by RVM can perform more accurate, which can reach 70% of accuracy.

According to Jana, et al (2013), they claimed that facial features can be used to estimate age group. Their experiment involved 3 stages: pre-processing, feature extraction and classification. During the pre-processing phase, they prepared datasets by taking images of 50 persons by using a digital camera (Nikon Coolpix L10). The face images were cropped, and the positions of eye pair, mouth, nose and chin were detected. During feature extraction, global features such as distances between 2 eye balls, eye to nose tip, eye to chin and eye to lip were determined. 6 types of ratios are then computed by referring to the distances obtained. After that, classification is carried out by using K-means clustering algorithm. Results showed ratio obtained using pixels (F5) was most reliable, with an accuracy of 96% when the samples are separated into 2 age groups, 84% of accuracy is obtained for 3 age groups and 4 age groups had accuracy of 62%.

2.3 DISCUSSION

From the above, the hence mentioned authors and their respective journals, they worked toward the objects estimation on the digital images. Some of them uses and implemented transfer learning that is pretrained models which includes Inception V3, GoogleNet and using SoftMax, many of them achieved accuracy over 90%. Some Authors proposed a related work to our study area by hierarchically method to estimate human age. They obtained the datasets in FG-Net website. They group the images in 3 major groups for training the model ranges (0-15), (15-30) and (30+). Their experiment also were performed in 3 steps, feature extraction, regression & classification.

CHAPTER THREE

SYSTEM ANALYSIS AND METHODOLOGY

3.1 INTRODUCTION

This chapter summarizes the procedures of doing research on the proposed system which is age and gender estimation methods. The main purpose of this work is to develop a framework of Age and gender prediction with deep convolutional neural network with OpenCV.

3.2 DESCRIPTION OF THE EXISTING SYSTEM

The existing system is a python program without GUI and it is not stating what is precisely the age and gender to a lay man. And in order to run it, you should have the knowledge of running it on the terminal. (Which means all the activities of training, testing and the evaluation of the program must be done in the in the terminal). Another drawback of the existing system is; every time we need to run it, the network needs to be retrained which will takes at least 2 hours of our time. That why we come up with training the network with the *Caffe model framework* which we train the CNN once and run it many times.

3.3 ANALYSIS OF THE EXISTING SYSTEM

3.3.1 EXISTING METHOD OF AGE CLASSIFICATION.

In recent years, many methods have been provided in the problem of automatically extracting age related attributes from facial images. A detailed survey of such methods of such methods has been presented in (Y. Fu et. al., 2010) and more recently in (H. Han et. Al, 2013).

Early methods for age estimation are based on calculating ratios between different measurement of facial features. Once a facial features (e.g. eye, nose, mouth, chin, etc.) are localized and their sizes and distances measured, ratios between them are calculated and for classifying the face into different age categories according to he hand-crafted rule(Y. H. Kwon et. al., 1994).

More recently, (N. Ramanathan and R. Chellappa, 2006) uses a similar approach to model age progression in subject under 18 years old. As the method require accurate localization (finding location of face feature such as nose) of facial features.

3.3.2 EXISTING METHOD OF GENDER CLASSIFICATION

A detailed survey of existing methods of gender classification has been provided in (Makinen and R. Raisamo, 2008), and more recently in (D. Reid et. al., 2013).

One of the early methods for gender classification used a neural network trained on a small set of near-frontal face images (B. A. Golomb et al., 1990).

A. J. O'toole et. al. uses the combined 3D structure of the head (obtained using a laser scanner) and image intensities were used for classifying gender (A. J. O'toole et al., 1997).

3.4 PROBLEMS OF THE EXISTING SYSTEM

Most of the methods discussed above used the FERET Dataset (P. J. Phillips et. al. both to develop the proposed systems and to evaluate performances. FERET images were taken under highly controlled condition. Moreover, the results obtained on this Dataset suggest that it is saturated and not challenging for modern methods. It is therefore difficult to estimate the actual relative benefit of these techniques. As a consequence, (C. Shan., 2012) experimented on the popular *Labeled Faces in the Wild (LFW) (G. B. Huang et al., 2007)* Dataset, primarily used for face recognition.

3.5 DESCRIPTION OF THE PROPOSE SYSTEM

As with age estimation, here too, we focus on the Adience dataset which contains images more challenging than those provided by *Labeled Faces in the Wild (LFW*) *(G. B. Huang et al., 2007)*, reporting performance using a more robust system, designed to better exploit information from massive example training sets.

3.6 ADVANTAGES OF THE PROPOSED SYSTEM OVER THE EXISTING SYSTEM

Some of the advantages of the proposed system includes the following:

- The design of the Graphical User Interface GUI for the proposed system
- It can be able to insert image and predict it with a click.
- It eliminates the burden of writing script on the terminal to estimate age and gender.

3.7 METHODOLOGY

The study begun with a review of related literature on the development of Age and Gender for developing the proposed system using image recognition technology. The literature research was accomplished through Google Scholar, Institute of Electrical and Electronics Engineers (IEEE). And the related works and researches done in recent years regarding the application of DL in image were studied in order to get an idea of developing the Age and Gender detection system. *Python* Programming language was chosen as software platform because of its flexibility and capability to support different DL packages.

3.7.1 DIFFERENCE BETWEEN NEURAL NETWORKS & DEEP NEURAL NETWORKS

In machine learning, a **convolutional neural network** (**CNN**, or **ConvNet**) is a class of deep,feed-forward artificial neural network that have successfully been applied to analyzing visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. They are also known as **shift invariant** or **space invariant artificial neural networks** (**SIANN**), based on their shared-weights architecture and translation invariance characteristics.

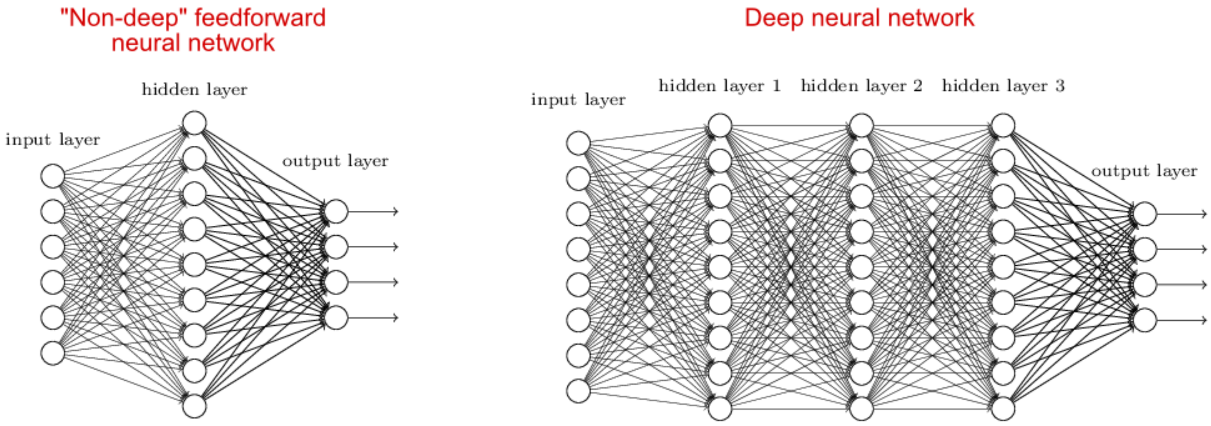

figure 3.1. Showing differences between non-deep NN and Deep NN

Convolutional networks were inspired by biological processes in which the connectivity pattern between neurons is inspired by the organization. Individual neurons respond to stimuli only in a restricted region of the visual field. CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage. They have applications in image and video recognition, recommender systems and natural language processing.

Deep learning is a class of machine learning algorithms that use a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. The algorithms may be supervised or unsupervised and applications include pattern analysis and classification are based on the learning of multiple levels of features or representations of the data, are part of the broader machine learning field of learning representations of data, learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.

These definitions have in common multiple layers of nonlinear processing units and the supervised or unsupervised learning of feature representations in each layer, with the layers forming a hierarchy from low-level to high-level features. The composition of a layer of nonlinear processing units used in a deep learning algorithm depends on the problem to be solved. Layers that have been used in deep learning include hidden layers of an artificial neural network and sets of complicated formulas.

At each layer, the signal is transformed by a processing unit, like an artificial neuron, whose parameters are iteratively adjusted through training.

3.7.2 NETWORK ARCHITECTURE

The network architecture is used throughout our experiments for both age and gender classification it is illustrated in Figure below. The network comprises of only three convolutional layers and two fully-connected layers with a small number of neurons. Age classification on the Adience set requires distinguishing between eight classes; gender only two. This, compared to, e.g., the ten thousand identity classes used to train the network used for face recognition in (Y. Sun. et al, 2014).

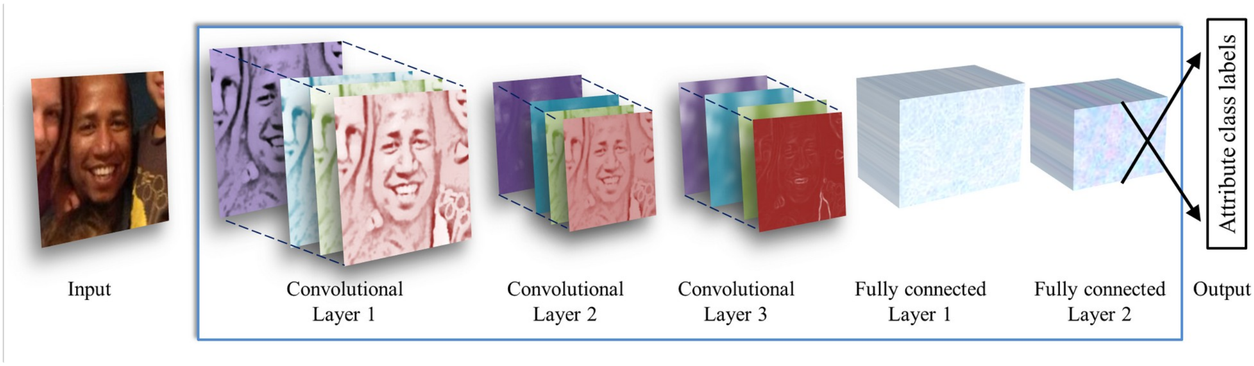

figure 3.2. *Illustration of our CNN architecture. The network contains three convolutional layers, each followed by a rectified linear operation and pooling layer. The first two layers also follow normalization using local response normalization (A. Krizhevsky et al., 2012). The first Convolutional Layer contains 96 filters of 7×7 pixels, the second Convolutional Layer contains 256 filters of 5×5 pixels, The third and final Convolutional Layer contains 384 filters of 3 × 3 pixels. Finally, two fully-connected layers are added, each containing 512 neurons. See Figure 2 for a detailed schematic view.*

All three color channels are processed directly by the network. Images are first re-scaled to 256×256 and a crop of 227×227 is fed to the network. The three subsequent convolutional layers are then defined as follows.

- 96 filters of size $3\times7\times7$ pixels are applied to the input in the first convolutional layer, followed by a rectified linear operator (ReLU), a max pooling layer taking the maximal value of 3×3 regions with two-pixel strides and a local response normalization layer *(A. Krizhevsky et al., 2012).*
- The $96 \times 28 \times 28$ output of the previous layer is then processed by the second convolutional layer, containing 256 filters of size $96 \times 5 \times 5$ pixels. Again, this is followed by ReLU, a max pooling layer and a local response normalization layer with the same hyper parameters as before.
- Finally, the third and last convolutional layer operates on the $256 \times 14 \times 14$ blob by applying a set of 384 filters of size $256 \times 3 \times 3$ pixels, followed by ReLU and a max pooling layer.

The following fully connected layers are then defined by:

- A first fully connected layer that receives the output of the third convolutional layer and contains 512 neurons, followed by a ReLU and a dropout layer.
- A second fully connected layer that receives the 512-dimensional output of the first fully connected layer and again contains 512 neurons, followed by a ReLU and a dropout layer.
- A third, fully connected layer which maps to the final classes for age or gender.

Finally, the output of the last fully connected layer is fed to a soft-max layer that assigns a probability for each class. The prediction itself is made by taking the class with the maximal probability for the given test image.

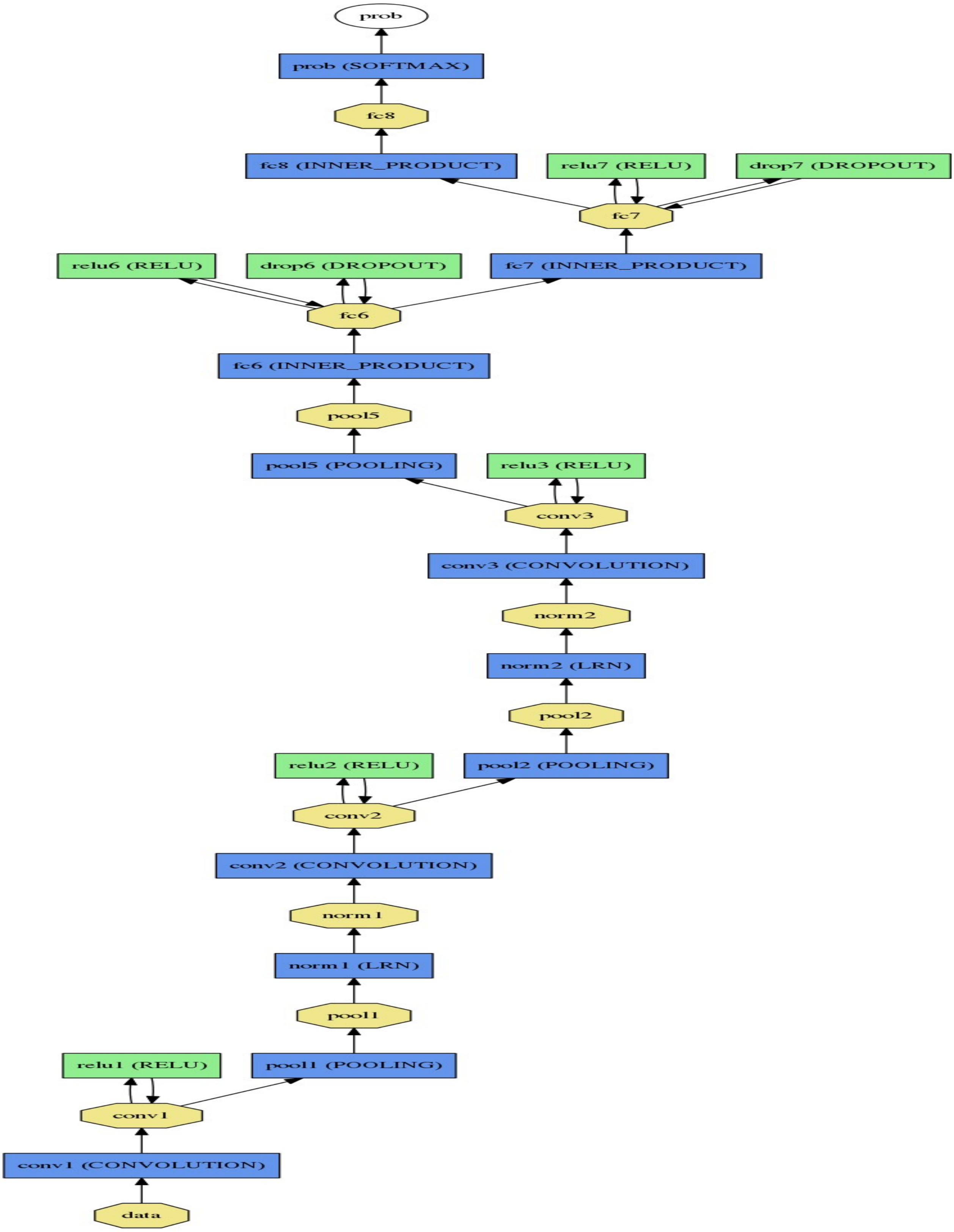

Figure 3.3 Full schematic diagram of our network architecture. Please see text under Figure 2. for more details.

3.7.3 PyQT5 FOR GUI DESIGN

Using Qt **Designer**. Qt **Designer** is the Qt tool for designing and building graphical user interfaces. It allows us to design widgets, dialogs or complete main windows using on-screen forms and a simple drag-and-drop interface.

PyQt is a python binding of the open-source widget-toolkit Qt, which also functions as a cross-platform application development framework. Qt is a popular C++ framework for writing GUI applications for all major desktop, mobile, and embedded platforms (supports Linux, Windows, MacOS, Android, iOS, Raspberry Pi, and more).

PyQt is developed and maintained by Riverbank Computing, a company based in England, whereas Qt is developed by a Finnish firm called The Qt Company.

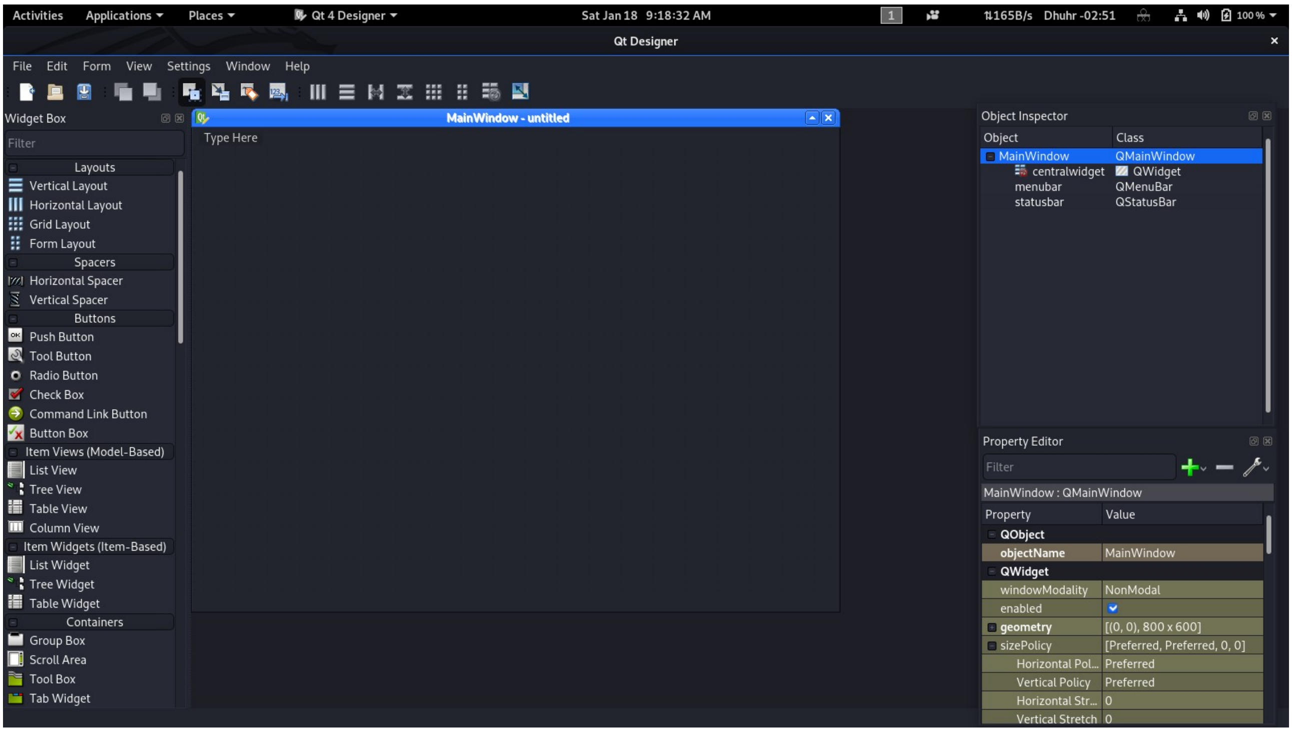

Figure 3.4 Qt Designer Main Window Home Screen

3.8 METHOD OF DATA COLLECTION

Data collection: Data collection plays an important role in training any deep neural network (DNN). In this project, the aim was to label data for two separate tasks: age and gender estimation. Collecting labeled data for some tasks, such as real age estimation, is much more challenging compared to popular classification or detection problems. This disparity is due to the fact that human error in estimating real age is large (sometimes greater than the computer vision estimations) and one cannot rely on human annotators to label faces with their corresponding real age. However, this project is done with a large dataset of faces with their corresponding age and gender labels. To our knowledge, this dataset is the largest or among the largest in either the academic or commercial world.

3.8.1 ADIENCE DATASE

Adience Database is large scale dataset, which was collected to capture all the variations in appearance, noise, pose, lighting and more, and can be expected to serve as images taken without careful preparation or posing. The Adience set consist of images automatically uploaded to flickr from smartphone devices. These images were uploaded without prior manual filtering. The entire Adience collection includes roughly 20k images of 2284 subjects. However, this database doesn’t have the accurate age annotation, i.e. age is annotated in a range form like 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+. (E. Eidinger et al, 2014).

	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60+	Total
Male	745	928	934	734	2308	1294	392	442	8192
Female	682	1234	1360	919	2509	1056	433	427	9411
Both	1427	2162	2294	1653	4897	2350	825	869	19487

Table 3.1 Adience image dataset distribution (number of images for each gender and age range)

3.8.2 DATASET PRE-PROCESSING

The dataset pre-processing step included converting human readable data, such as images into computer readable binary data. In this project, the images were converted to **Caffe Model** format data as the input for training and testing process.

3.9 CAFFE MODEL

Caffe (Convolutional Architecture for Fast Feature Embedding) model is a deep learning framework made with expression, speed, and modularity in mind. It is developed by Berkeley AI Research (BAIR) and by community contributors. We choose caffe in this project because of the following:

1. **Speed**: makes Caffe perfect for research experiments and industry deployment. Caffe can process over 60M images per day with a single NVIDIA K40 GPU*. That's 1 ms/image for inference and 4 ms/image for learning and more recent library versions and hardware are faster still. We believe that Caffe is among the fastest ConvNet implementations available.
2. **Expressive architecture**: Which encourages application and innovation. Models and optimization are defined by configuration without hard-coding.
3. **Extensible code:** fosters active development. In Caffe's first year, it has been forked by over 1,000 developers on github and had many significant changes contributed back.
4. **Community:** Caffe already powers academic research projects, startup prototypes, and even large-scale industrial applications in vision, speech, and multimedia.

This project method is implemented using the Caffe open-source framework the training was performed based on (Y. Jia. Et al., 2014). Prediction running times can conceivably be substantially improved by running the network on image batches.

CHAPTER FOUR

SYSTEM DESIGN AND IMPLEMENTATION

4.1 INTRODUCTION

In this chapter, I highlighted some of the system algorithm and pseudocode. User interfaces and also the tools that help in the implementation of the system. And discussion followed at the end.

4.2 SYSTEM DESIGN

The design phase of any system is very important, vital and crucial because the success of any system depends largely on its design specifications. In this phase, the final specifications are used for translating the model into a design of the desired system and modules are being defined showing their relationships to one another in a way known as a structural chart using structured tools. The reason for the design phase is to specify a particular software system that will meet the requirements gathered at the analysis phase. Structured design divides a program into smaller, independent modules. They are arranged orderly in a hierarchy that shows a model of the application area which is organized in a top-down manner.

The concept of modification thus comes from structured design which is an attempt to reduce complexity and make a problem manageable by sub-dividing it into smaller segments.

4.3 PHYSICAL DESIGN

a) Input Design

This is an interface between the user and the system that allows the user to enter data. Image input is generally done through either the webcam or file upload.

b) Output Design

This serves as an interface between the user and the system that provide report to the user.

4.4 APPLICATION ALGORITHM

In mathematics and computer science, an algorithm is an effective method expressed as a finite list of well-defined instructions for calculating a function. Algorithms are used for calculation, data processing, and automated reasoning.

Starting from an initial state and initial input (perhaps null), the instructions describe a computation that, when executed, will proceed through a finite number of well-defined successive states, eventually producing "output" and terminating at a final ending state. The transition from one state to the next is not necessarily deterministic; some algorithms, known as randomized algorithms, incorporate random input.

step 1: Open the webcam or accept the image as a file.

Step 2: Restart/Start reading time.

Step 3: Read image frame from webcam

Step 4: Use OpenCV library to find the face in the frame and if there is no face detected go to Step 2.

Step 5: Use ageNet Convolutional Neural Networks (forward) to estimate the age.

Step 6: Use genderNet Convolutional Neural Networks (forward) to estimate the Gender.

Step 7: Display age and Gender.

Step 8: Display accuracy of both age and Gender.

Step 9: Display time taken to estimate age and Gender.

Step 10: Go to Step 1 for another frame or quit.

4.4.1 PSEUDOCODE

```
while camera is open or image file imported
     start timer
     get frame

     if no frame
          break
     get box over the subject face
     if no box detected  //meaning that no face detected
          continue
     for every face in the box
          extract face features
          insert the face features as input of caffemodel

          //for gender
          use caffemodel to estimate the gender of face
          display estimated gender
          display the accuracy of the estimated gender

          //for age
          use caffemodel to estimate the age of face
          display estimated age
          display the accuracy of the estimated age

          display time taken to estimate
```

```
if webcam is open
      continue
else
      break
```

4.5 USER INTERFACE

4.5.1 SPLASH SCREEN

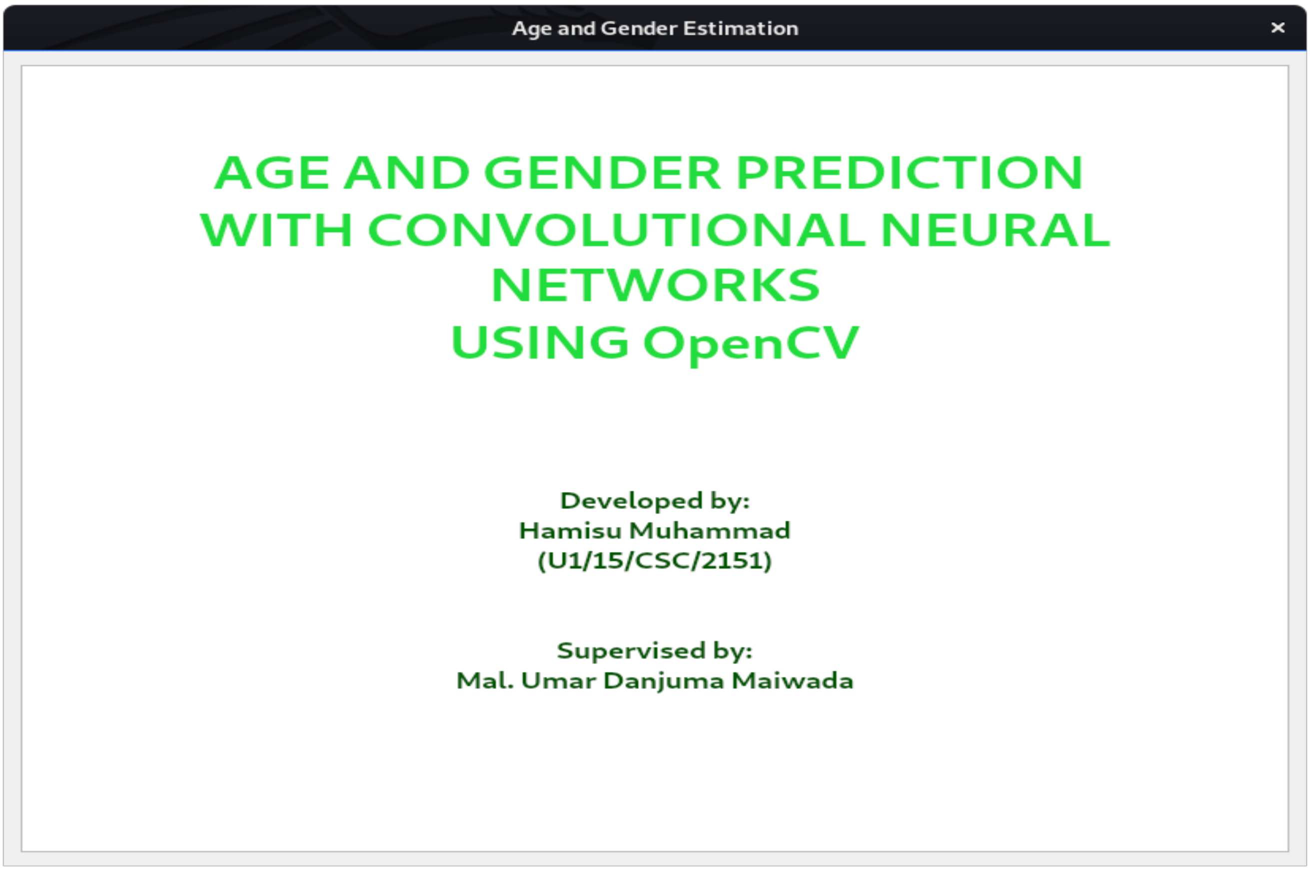

Figure 4.1 Splash screen for stating the system

4.5.2 MAIN MENU

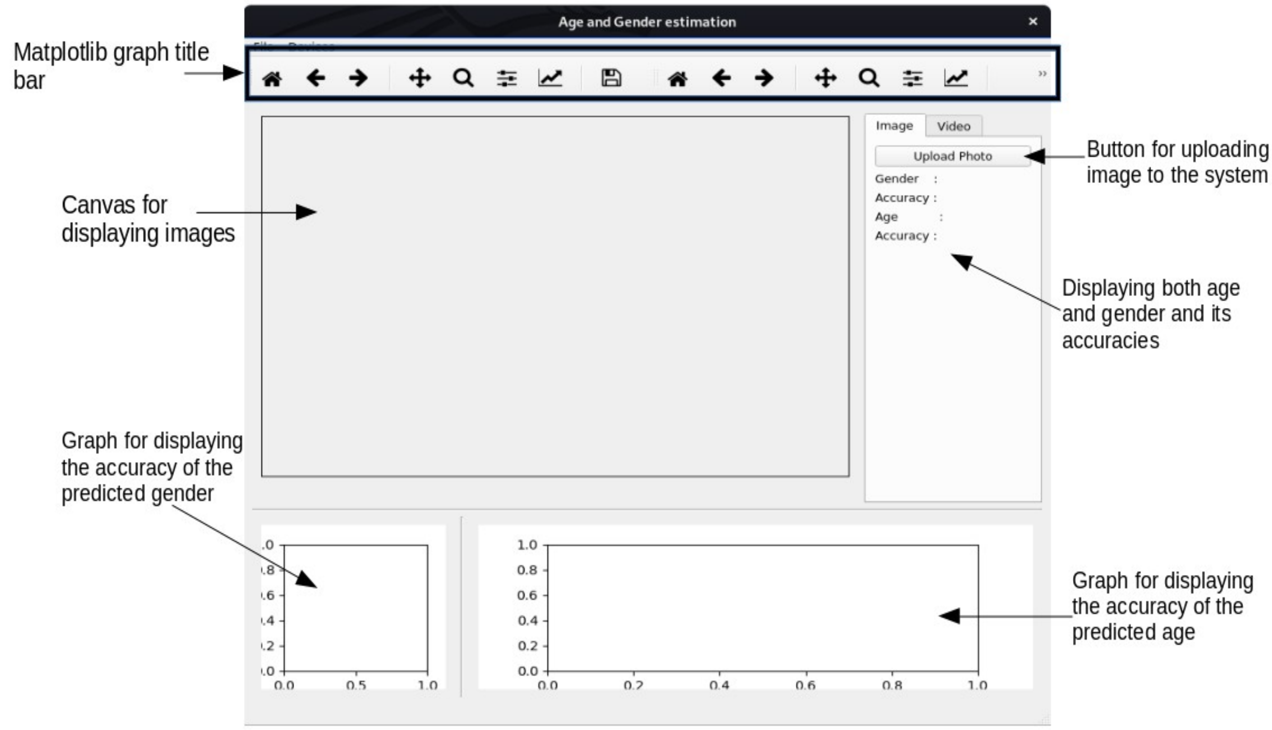

Figure 4.2 system menu for image estimation

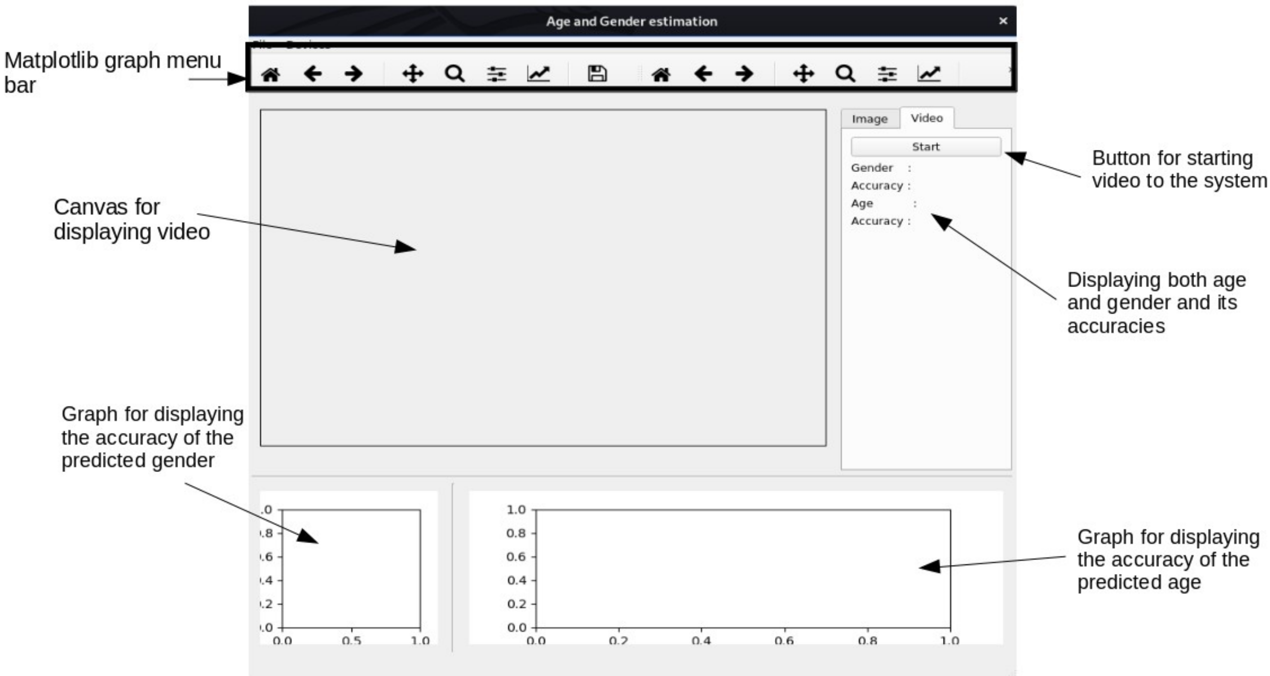

Figure. 4.3 system menu for real-time estimation

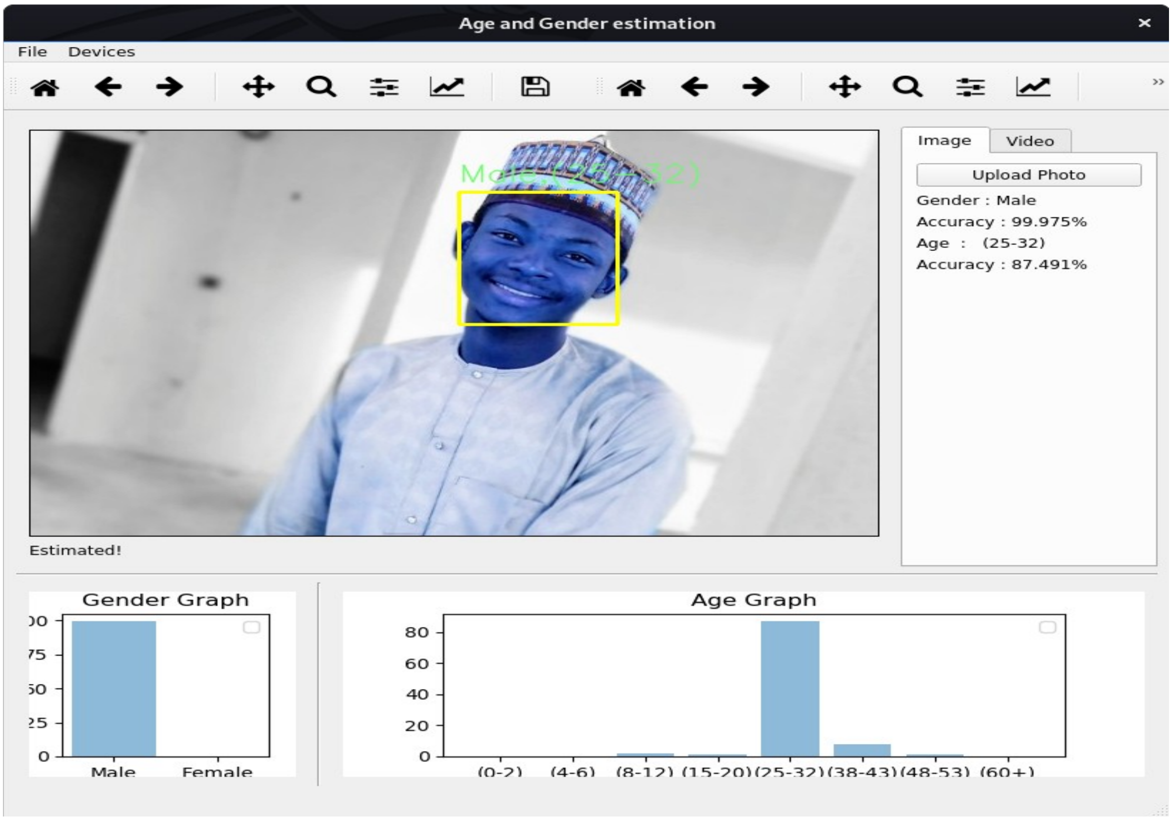

Figure 4.4 Testing the system with image

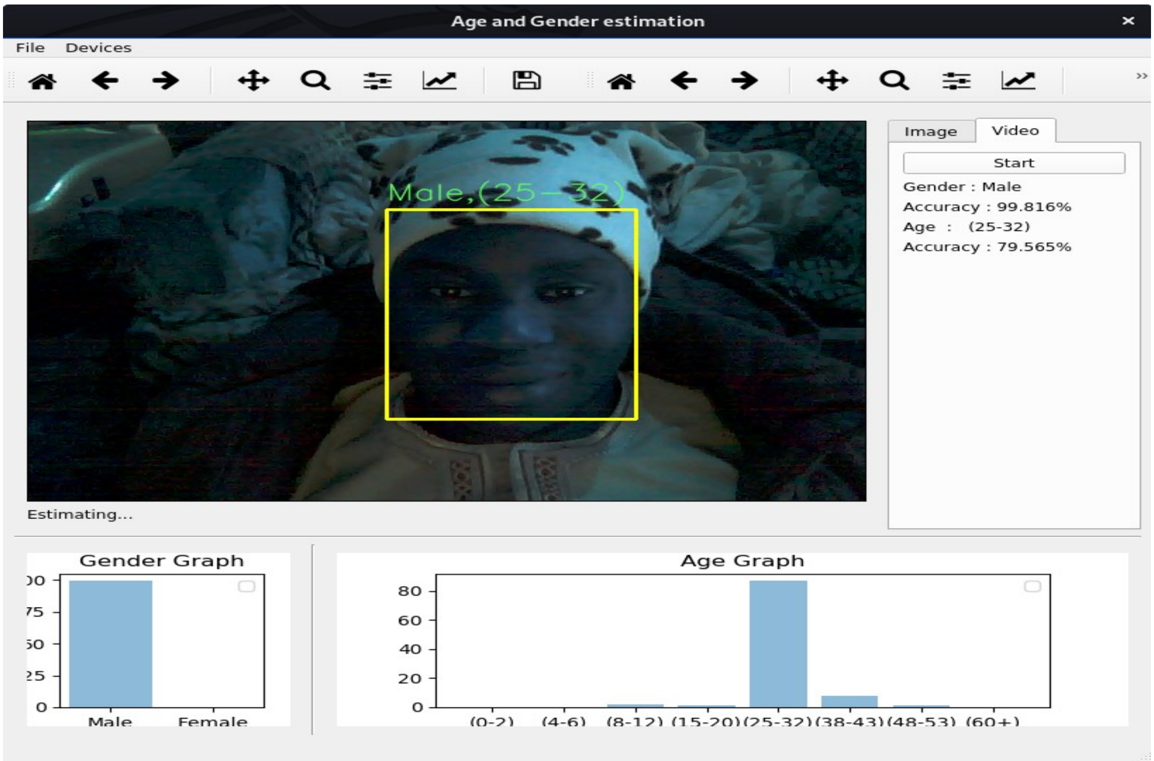

Figure 4.5 Real-time estimation

4.6 SYSTEM REQUIREMENTS

These are set of tools needed by the system to operate, as its design. The proper operation of the newly designed system depends on these requirements. These requirements are in two classes, hardware and software requirement.

- **Minimum Hardware requirement**

This is the physical component needed by the system to operate. The software require moderate hardware to manage multiple processing. This software requires a minimum of:

- Build-in camera of a laptop (used as image input device)
- Pentium processor 2.27 GHz
- 4 GB of RAM
- 100 GB of storage

- **Minimum Software requirement**

This is the non-physical component needed by the system to make it operable.

- Windows operating systems: Windows 7, windows 8, windows 10 or latter version with Python and other dependencies installed.
- Linux Operating System: Ubuntu, Parrot, Kali, etc.

4.7 IMPLEMENTATION TOOLS

Below are some of the tools used to run the project design, which includes:

- **Python Programming Language**

Python is a programming language with extensive supported packages and modules. It is developed by Guido van Rossum. It is derived from many other languages such as ABC, modula-3, C, C++, Algol-68, smallTalk, Unix shell and other scripting languages. Python also provides interfaces to all major commercial databases (Swaroop, 2003). Python consists of a broad standard library. This feature enables the exploration and access to various file types such as XML, HTML, WAV, CSV files.

IDLE is a simple Python Integrated Development Environment (IDE) available for Windows, Linux, and Mac OS X (Lent, 2013). All commands can be typed, saved and run in Python IDLE iterative shell. As such, Python is chosen as programming language throughout the project.

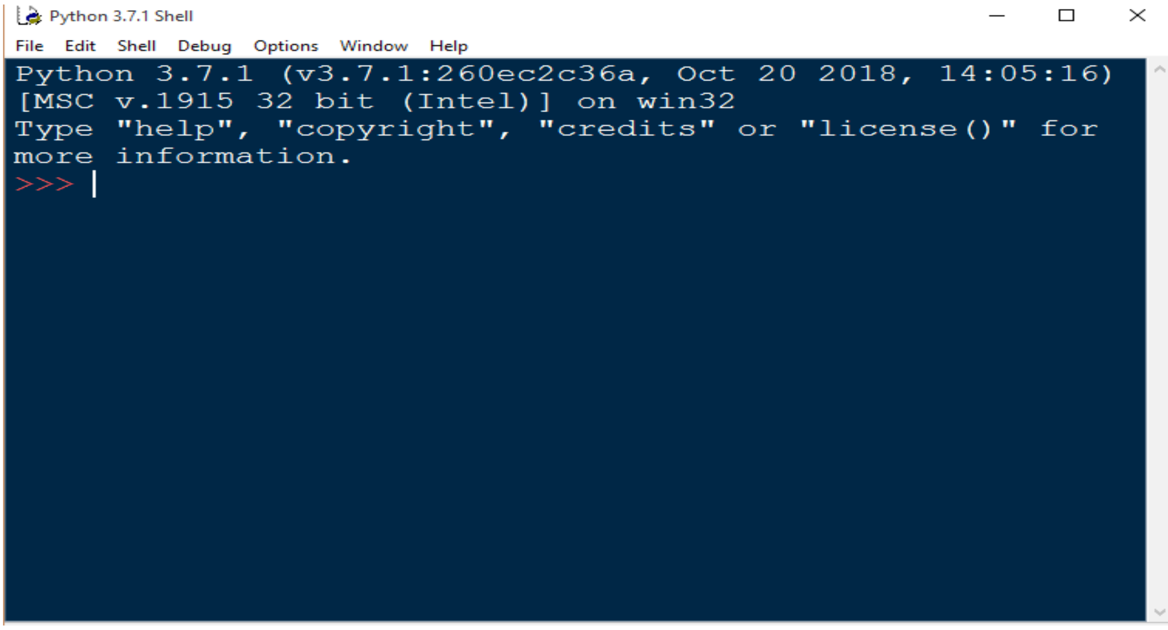

Figure 4.6 Python IDLE (Python shell)

- **Anaconda (Python) Distribution**

Anaconda is a free and open-source distribution of the *Python* and *R* programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. Package versions are managed by the package management system *conda*. The Anaconda distribution is used by over 15 million users and includes more than 1500 popular data-science packages suitable for Windows, Linux, and MacOS.

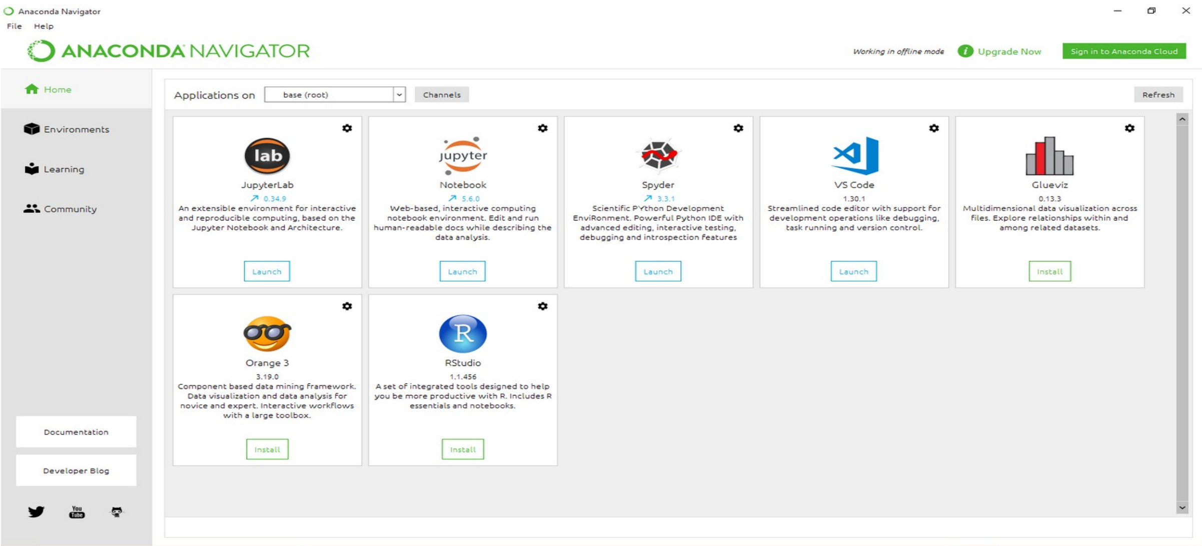

Figure 4.7 Anaconda Navigator Home Menu

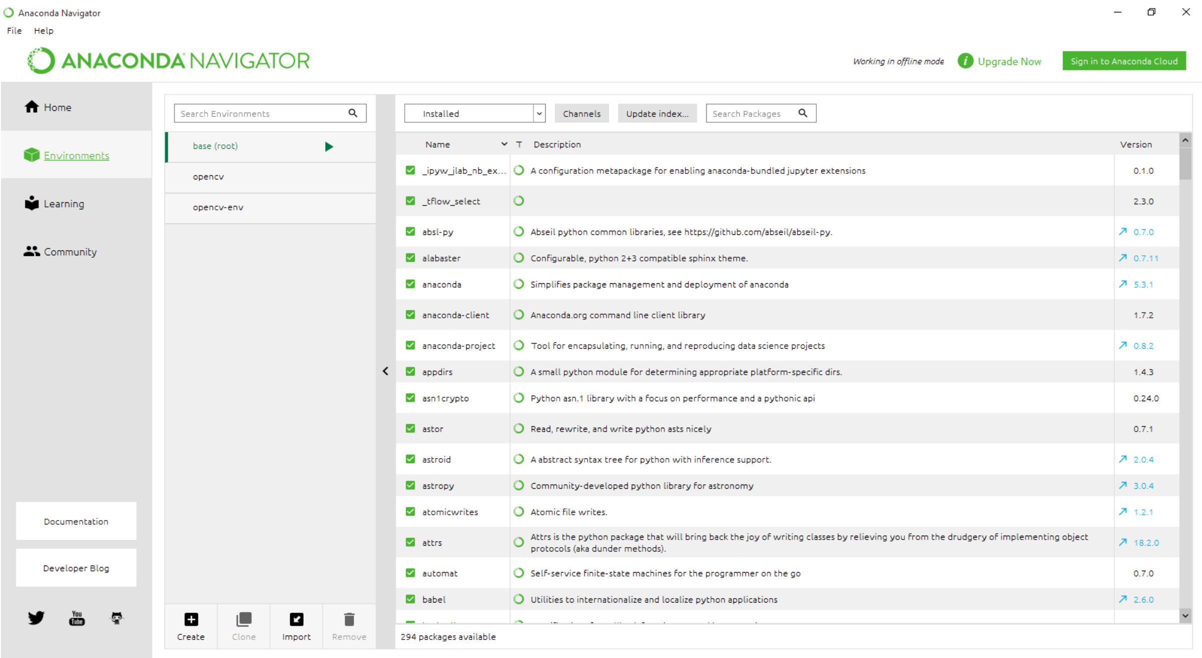

Figure 4.8 Anaconda Navigator List of Packages

- **Open Source Computer Vision Library (OpenCV)**

OpenCV (Open Source Computer Vision Library) is a library of programming functions mainly aimed at real time computer vision, developed by Intel. The library is cross-platform. It focuses mainly on real-time image processing. Python, Ruby and Java (using JavaCV) have been developed to encourage adoption by a wider audience.

However, since version 2.0, OpenCV includes both its traditional C interface as well as a new C++ interface. This new interface seeks to reduce the number of lines of code necessary to code up vision functionality as well as reduce common programming errors such as memory leaks (through automatic data allocation and deallocation) that can arise when using OpenCV in C. Most of the new developments and algorithms in OpenCV are now developed in the C++ interface. Unfortunately, it is much more difficult to provide wrappers in other languages to C++ code as opposed to C code; therefore the other language wrappers are generally lacking some of the newer OpenCV 2.0 features.

CHAPTER FIVE

SUMMARY, RECOMMENDATIONS AND CONCLUSION

5.1 SUMMARY

In this project we have gone through the design and developing a system that estimate the age and gender of a person in the given image and in real-time using webcam as video input device. Due to the increasing interest in social robotics and video-based security systems, research on the numerical analysis of human faces (including face detection, face recognition, classification of gender, and recognition of facial expression) has attracted attention in the communities of computer vision and pattern recognition. In connection with these investigations, estimating the age and gender of a person from the numerical analysis of his/her face image is a relatively new topic. Age estimation by numerical analysis of the face image has many potential applications such as the development of intelligent human-machine interfaces and improving safety and protection of minors in various and diverse sectors (transport, medicine, etc.). It can be very useful for advanced video surveillance, demographic statistics collection, business intelligence and customer profiling, and search optimization in large databases. The age and gender attribute could also be used in the verification of the face and enriching the tools used in police investigations. In general, automatic age and gender estimation by a machine is useful in applications where the objective is to determine the age and gender of an individual without identifying him.

5.2 RECOMMENDATION

A further improvement is needed to advance the system in order to adapt with real life application. The precision of the system needs to improve in order to produce a more reliable device. This can be done by increasing the quality of training dataset. One can also try to use a regression model instead of classification for Age Prediction if enough data is available.

5.3 CONCLUSION

Convolutional Neural Network (CNN) can be used to provide improved age and gender classification results, even considering the much smaller size of contemporary unconstrained image sets labelled for age and gender. The simplicity of the model implies that more elaborate systems using more training data may well be capable of substantially improving results beyond these results obtained in this project.

References

- Dietterich, T., & Michalski, R. S. (1983). A Comparative Review of Selected Methods for Learning From Examples. *Machine Learning: An Artificial Intelligence Approach*. https://doi.org/10.1007/978-3-662-12405-5
- Luppescu, G., & Romero, F. (n.d.). Comparing Deep Learning and Conventional Machine Learning for Authorship Attribution and Text Generation, 1 –9.
- Negnevitsky, M. (2002). *Artificial Intelligence: A Guide to Intelligent Systems* (2nd ed.). Harlow, Essex: Pearson Education Limited
- Michie, D., Spiegelhalter, D. J., Taylor, C. C., Michie, E. D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine Learning, Neural and Statistical Classification. *Ellis Horwood Series in Artificial Intelligence*, *37*(4), xiv, 289 . https://doi.org/10.2307/1269742
- Weinberger, K., Blitzer, J., & Saul, L. (2006). Distance metric learning for large margin nearest neighbor classification. *Advances in Neural Information Processing Systems*, *18*, 1473. https://doi.org/10.1007/978-3-319-13168-9_33
- Tan, P.N., Steinbach, M., & Kumar, V. (2005a). Chap 8 : Cluster Analysis: Basic Concepts and Algorithms. *Introduction to Data Mining*, Chapter 8. https://doi.org/10.1016/0022-4405(81)90007-8
- Tan, P. N., Steinbach, M., & Kumar, V. (2005b). Association Analysis: Basic Concepts and Algorithms. *Introduction to Data Mining*, 327–414. https://doi.org/10.1111/j.1600-0765.2011.01426.x
- Mnih, V., Silver, D., & Riedmiller, M. (n.d.). Dqn. *Nips*, 1–9. https://doi.org/10.1038/nature14236
- Barto, A., & Dietterich, T. (2004). Reinforcement learning and its relationship to supervised learning. *Handbook of Learning and Approximate Dynamic Programming*, 47–64. https://doi.org/10.1002/9780470544785.ch2
- Jain, A. K., & Mao, J. (1996). Artificial Neural Network: A Tutorial. *Communications*, *29*, 31–44. https://doi.org/10.1109/2.485891
- Popescu, M. C., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. (2009). Multilayer Perceptron and Neural Networks. *WSEAS Transactions on Circuits and Systems*, *8*(7), 579–588.

- Popescu, M. C., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. (2009). Multilayer Perceptron and Neural Networks. *WSEAS Transactions on Circuits and Systems*, *8*(7), 579–588.

- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

- Zhong, J., Peniak, M., Tani, J., Ogata, T., & Cangelosi, A. (2016). Sensorimotor Input as a Language Generalisation Tool: A Neurorobotics Model for Generation and Generalisation of Noun-Verb Combinations with Sensorimotor Inputs, (May), 1 – 23. Retrieved from http://arxiv.org/abs/1605.03261
- Nielsen, M. A. (2015). Neural Networks and Deep Learning. Retrieved 2018, from http://neuralnetworksanddeeplearning.com/chap5.html
- Zhong, J., Peniak, M., Tani, J., Ogata, T., & Cangelosi, A. (2016). Sensorimotor Input as a Language Generalisation Tool: A Neurorobotics Model for Generation and Generalisation of Noun-Verb Combinations with Sensorimotor Inputs, (May), 1 – 23. Retrieved from http://arxiv.org/abs/1605.03261
- Devikar, P. (2016). Transfer Learning for Image Classification of various dog breeds. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *5*(12).
- Chin, C. S., Si, J., Clare, A. S., & Ma, M. (2017). Intelligent Image Recognition System for Marine Fouling Using Softmax Transfer Learning and Deep Convolutional Neural Networks. *Complexity*, *2017*. https://doi.org/10.1155/2017/5730419
- Tamkin, A., & Usiri, I. (2013). Deep CNNs for Diabetic Retinopathy Detection, 1 –6.
- Thukral, P., Mitra, K., & Chellappa, R. (2012). A hierarchical approach for human age estimation. *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, (March), 1529–1532. https://doi.org/10.1109/ICASSP.2012.6288182
- Jana, R., Datta, D., & Saha, R. (2013). Age Group Estimation using Face Features, *3*(2), 130–134.

- Craw, I., Ellis, H., and Lishman, J.R. (1987). Automatic extraction of face features. Pattern Recognition Letters, 5:183-187, February.
- W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. volume 35, pages 399{458, New York, NY, USA, December 2003. ACM.
- Rabia Jafri and Hamid R. Arabnia. A survey of face recognition techniques. volume 5, pages 41{68, 2009.
- Matthew Turk and Alex Pentland. Eigenfaces for recognition. volume 3, pages 71{86, Cambridge, MA, USA, 1991. MIT Press.
- S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller. Fisher discriminant analysis with kernels. In Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop., pages 41-48, Aug 1999.
- Karl Pearson. On lines and planes of closest fit to systems of points in space. volume 6, pages 559-572, 1901.
- H. Hotelling. Analysis of a complex of statistical variables into principal components. 1933.

- S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller. Fisher discriminant analysis with kernels. In Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop., pages 41-48, Aug 1999.
- Kamran Etemad and Rama Chellappa. Discriminant analysis for recognition of human face images (invited paper). In Josef Bign, Grard Chollet, and Gunilla Borgefors, editors, AVBPA, volume 1206 of Lecture Notes in Computer Science, pages 127-142. Springer, 1997.
- Choon Boon Ng, Yong Haur Tay, and Bok-Min Goi. Vision-based human gender recognition: A survey. volume abs/1204.1611, 2012.
- Erno Mkinen and Roope Raisamo. An experimental comparison of gender classification methods. volume 29, pages 1544 - 1556, 2008.
- Prajakta A. Mélange, Dr. G. S. Sable. “Age Group Estimation and Gender Recognition Using Face Features” The International Journal of Engineering and Science (IJES) Volume 7 Issue 7 Ver. I Pages PP 01-07, 2018 ISSN (e): 2319 – 1813 ISSN (p): 23-19 – 1805.
- W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. volume 35, pages 399-458, New York, NY, USA, December 2003.ACM.
- Eran Eidinger, Roee Enbar, and Tal Hassner. Age and gender estimation of unfiltered faces. IEEE Transactions on Information Forensics and Security, 9(12):2170–2179, 2014.