Gender and Age Detection

Department of Computer Science and Engineering

Bangladesh University of Business and Technology

Dhaka-1216

November 2021

Gender and Age Detection

A report

submitted to the department of Computer Science and Engineering
in partial fulfillment of the requirements
for the degree of

Bachelor of Science in Computer Science and Engineering

By

Shakil Ahmed (ID:17182103010)

Mahmudul Hasan (ID:18191103057)

Mahady hasan (ID:18191103061)

Momin Ali (ID: 38-17182103011)

MD Tanvir Hossain Reza (ID:38-17182103020)

Supervised by

MD. Shahiduzzaman

Assistant Professor

Department of Computer Science and Engineering (CSE)
Bangladesh University of Business and Technology (BUBT)
Mirpur-2, Dhaka-1216, Bangladesh

ABSTRACT

Nowadays research has explored extracting auxiliary information from various biometric techniques such as fingerprints, face, iris, palm, voice etc. This information contains some features like gender, age, beard, moustache, scars, height, hair, skin color, glasses, weight, facial marks, tattoos etc. All this information contributes more and more during identification. The major changes that come across face recognition is to find the age & gender of the person. This paper contributes a significant survey of various face recognition techniques for finding the age and gender. The existing techniques are discussed based on their performances. This paper also provides future directions for further research. Age and gender, two of the key facial attributes, play a very foundational role in social interactions, making age and gender estimation from a single face image an important task in intelligent applications, such as access control, human-computer interaction, law enforcement, marketing intelligence and visual surveillance.

This project is based upon computer vision and the various terminologies used to process images and detect age and gender of the person from the image.

A Convolutional Neural work is a deep neural network (DNN) widely used for the purposes of image recognition and processing and NLP. Also known as a ConvNet, a CNN has input and output layers, and multiple hidden layers, many of which are convolutional. In a way, CNNs are regularized multilayer perceptrons. Computer Vision is the field of study that enables computers to see and identify digital images and videos as a human would. The challenges it faces largely follow from the limited understanding of biological vision. A fast and efficient gender and age estimation system based on facial images is developed. There are many methods that have been proposed in the literature for the age estimation and gender classification. However, all of them have still disadvantage such as not complete reflection about face structure, face texture. This technique applies to both face alignment and recognition and significantly improves three aspects Within a given database, all weight vectors of the persons within the same age group are averaged together. A range of an age estimation result is 15 to 70 years old, and divided into 13 classes with 5 years old range. Experimental results show that better gender classification and age estimation.

age and gender classification has become applicable to an extending measure of applications, particularly resulting to the ascent of social platforms and social media. Regardless, execution of existing strategies on real-world images is still fundamentally missing, especially when considering the immense bounce in execution starting late reported for the related task of face acknowledgment. In this paper we exhibit that by learning representations through the use of significant Convolutional Neural Systems (CNN), a huge augmentation in execution can be acquired on these errands. To this end, we propose a direct Convolutional Neural System engineering can be used despite when the measure of learning data is limited. We survey our procedure on the recent Adience benchmark for age and gender estimation and demonstrate it to radically outflank current state of-the-art methods.

DECLARATION

We declare that this thesis and the work presented in it are our own and has been generated by us as the result of our own original research

We confirm that:

- This Work is done wholly or mainly while in candidature for a research degree at this University.
- This report work has not been previously submitted for any degree at this university or any other educational institutes.
- We have quoted from the work of others; the source is always given. With the exception of such quotations, this thesis is entirely our own work.

Shakil Ahmed
ID: 17182103010
Intake: 38

Mahmudul Hasan
ID: 18191103057
Intake: 40

Mahady Hasan
ID: 18191103061
Intake: 40

MD Tanvir Hossain Reza ID: 17182103020
Intake: 38

Momin Ali
ID: 171812103011
Intake: 38

CERTIFICATE

This is to certify that **Shakil Ahmed, Mahmudul Hasan, Mahady Hasan, Mohammad Tanvir Hossain Reza and Momin Ali** students of B.Sc. in CSE have completed their report work titled "Age and Gender Detection" satisfactorily in partial fulfillment for the requirement of B.Sc.in CSE. Bangladesh University of Business and Technology in the year 2021.

Shakil Ahmed
ID: 17182103010
Intake: 38

Mahmudul Hasan
ID: 18191103057
Intake: 40

Mahady Hasan
ID: 18191103061
Intake: 40

Mohammad Tanvir Hossain Reza
ID: 17182103020
Intake: 38

Momin Ali
ID: 171812103011
Intake: 38

Project Supervisor

(MD. Shahiduzzaman)

Assistant Professor

Department of Computer Science and Engineering (CSE)
Bangladesh University of Business and Technology (BUBT)

DEDICATION

Dedicated to our parents for all their love and inspiration.

ACKNOWLEDGEMENTS

First of all, we are thankful and expressing our gratitude to Almighty Allah who offers us His divine blessing, patience, mental and physical strength to complete this project work.

We are deeply indebted to our project supervisor MD. Shahiduzzaman, Assistant Professor, Department of Computer Science and Engineering (CSE), Bangladesh University of Business and Technology (BUBT). His scholarly guidance, important suggestions, work for going through our drafts and correcting them, and generating courage from the beginning to the end of the research work has made the completion of this project possible.

A very special gratitude goes out to all our friends for their support and help to implement our works. The discussions with them on various topics of our works have been very helpful for us to enrich our knowledge and conception regarding the work.

Last but not the least; we are highly grateful to our parents and family members for supporting us spiritually throughout writing this report and our life in general.

APPROVAL

This project report "**Age and Gender Detection"** Submitted by Shakil Ahmed, Mahmudul Hasan, Mahady Hasan, Mohammad Tanvir Hossain Reza and Momin Ali ID NO: 17182103010, 18191103057, 18191103061, 17182103020, 17182103011 Department of Computer Science and Engineering (CSE), Bangladesh University of Business and Technology (BUBT) under the supervision of MD. Shahiduzzaman, Assistant Professor, Department of Computer Science and Engineering has been accepted as satisfactory for the partial fulfillment of the requirement for the degree of Bachelor of Science (B.Sc. Eng.) in Computer Science and Engineering and approved as to its style and contents.

Supervisor:

MD. Shahiduzzaman

Assistant Professor
Department of Computer Science and Engineering (CSE)
Bangladesh University of Business and Technology (BUBT)
Mirpur-2, Dhaka-1216, Bangladesh

Chairman:

Dr.Muhammad Firoz Mridha

Associate Professor & Chairman
Department of Computer Science and Engineering (CSE)
Bangladesh University of Business and Technology (BUBT)
Mirpur-2, Dhaka-1216, Bangladesh

Abbreviations

Synonyms and Acronyms	Descriptions
CNN	Convolutional Neural Network
ANN	Artificial Neural Network

Table of Content

List of Figures

List of Tables

N/A

Chapter 01

Introduction

3.4 Introduction

Age and gender, two of the key facial attributes, play a very foundational role in social interactions, making age and gender estimation from a single face image an important task in intelligent applications, such as access control, human-computer interaction, law enforcement, marketing intelligence and visual surveillance, etc. The enhancement of raw images that are received from the camera sources, from satellites, aircrafts and the pictures captured in day-to-day lives is called image processing. The images have been processed through many different techniques and calculations have been made

on the basis and analysis of the studies. There is a need to analyze and study the digitally formed images. There are two main and very common steps followed for image processing. The improvement of an image such that the resulting image is of greater quality and can be used by other programs, is called image enhancement. The other technique is the most sought-after technique used for extraction of information from an image. There is a division of the image into a certain number of parts or objects so that the problem is solved. This process is called segmentation. A neural network consists of many simple and similar compressing elements. It is a system with inputs and outputs. There are a number of internal parameters called weights. An artificial neural network is made of a set of processing elements which are also known as neurons or nodes. These nodes are interconnected. Training in ANN is done through the track of the examples. There are various such methods that fail to produce appropriate results. For each class, an essential rule called the characteristic rule is generated. This set of rules is also called differentiating rules. A systematic method which is used to train multilayer artificial neural networks is known as back propagation. It is also considered as a gradient method where the

gradient of the error is evaluated by considering the weights of the given inputs. The detection of the data available in the images is very important. The data that the image contains is to be changed and modified for the detection purposes. There are various types of techniques involved for detection as well as the removal of the problem. In a Facial detection technique: The expressions that the faces contain hold a lot of information. Whenever a person interacts with the other person, there is an involvement of a lot of expressions. The changing of expressions helps in calculating certain parameters. Age estimation is a multi-class problem in which the years are classified into classes. People with different ages have different facials, so it is difficult to gather the images. Various age detection methods are used. The preprocessing is applied to the image. Features are extracted from the neural network through the convolution network. Based on the trained models the image is then classified to one of the age classes. Features are extracted from the images for further processing. The features are processed further and sent to the training systems. The databases provide a study of the features and help in completing the face detection for proving the age detection of the person in the image. Age and gender assume essential parts in social activities. Dialects hold distinctive greetings and grammar rules for men or women, and frequently diverse vocabularies are utilized while tending to senior citizens compared to youngsters. In spite of the essential parts these characteristics play in our everyday lives, the capacity to consequently assess them precisely and dependably from face image is still a long way from addressing the requirements of business applications. This is especially puzzling while considering late claims to super-human capacities in the related errand of face recognition. .Past ways to deal with assessing or ordering these properties from face images have depended on contrasts in facial feature dimensions or “customized” face descriptors. Most have utilized characterization plans composed especially for age or gender orientation estimation undertakings, including and others. Few of these past strategies were intended to handle the numerous difficulties of unconstrained imaging conditions . In addition, the machine learning strategies utilized by these frameworks did not completely abuse the huge quantities of image cases and information accessible through the Internet keeping in mind the end goal to enhance characterization capacities.In this paper we endeavour to close the gap between automatic face recognition abilities and those of age and

gender classification techniques. To this end, we take after the fruitful sample set around late face recognition frameworks: Face recognition systems portrayed in the most recent couple of years have demonstrated that gigantic advancement can be made by the utilization of profound convolutional neural networks (CNN) . We show comparative additions with basic system engineering, composed by considering the somewhat constrained accessibility of precise age and gender classification names in existing face information sets.

1.2 Problem Statement

Age and gender characteristics can be applied in retail for contextual advertising for particular groups of customers. The reliability of the existing solutions remains insufficient for practical application. Developing an age and gender recognition using convolutional neural networks with classifier fusion methods

- To conduct a review of the existing solutions for image recognition by age and gender with an emphasis on convolutional neural networks (CNNs)
- To describe the algorithm scheme of the proposed system
- To develop the age/gender recognition architecture, taking into account the classifier fusion methods
- To analyze the accuracy of decision making on the basis of each aggregation method by conducting experiments

1.3 Problem Background

Different methods have been proposed for extracting features from face images and to train the system for identifying gender. We have experimented on different feature extraction methods and been adopted for our proposed system. In the following subsections we will briefly describe some necessary topics related to our proposed system.

1.4 Research Objectives

The objective of this study is as follows research:

- To Identify Age of Human
- To Identify Gender of Human
- To Achieve Better Accuracy

1.5 Motivation

The motivation behind this project was to build an application for age and gender classification using a model that provides better accuracy. Many models are focusing on real life detection in-the-wild estimation. That doesn't provide better accuracy. In this project we will focus on CNN methods. That will help us to achieve good accuracy.

1.6 Flow of Research

The research work is being developed in several steps. Firstly we analyzed the research topic then studied the basic theory of CNN and OpenCV. Then we investigated the application of CNN and OpenCV. We investigated the lack of present architectures and motivated them to build a new architecture based on CNN architecture. Figure 1.1 illustrates the overall steps to the research procedure in the following diagram.

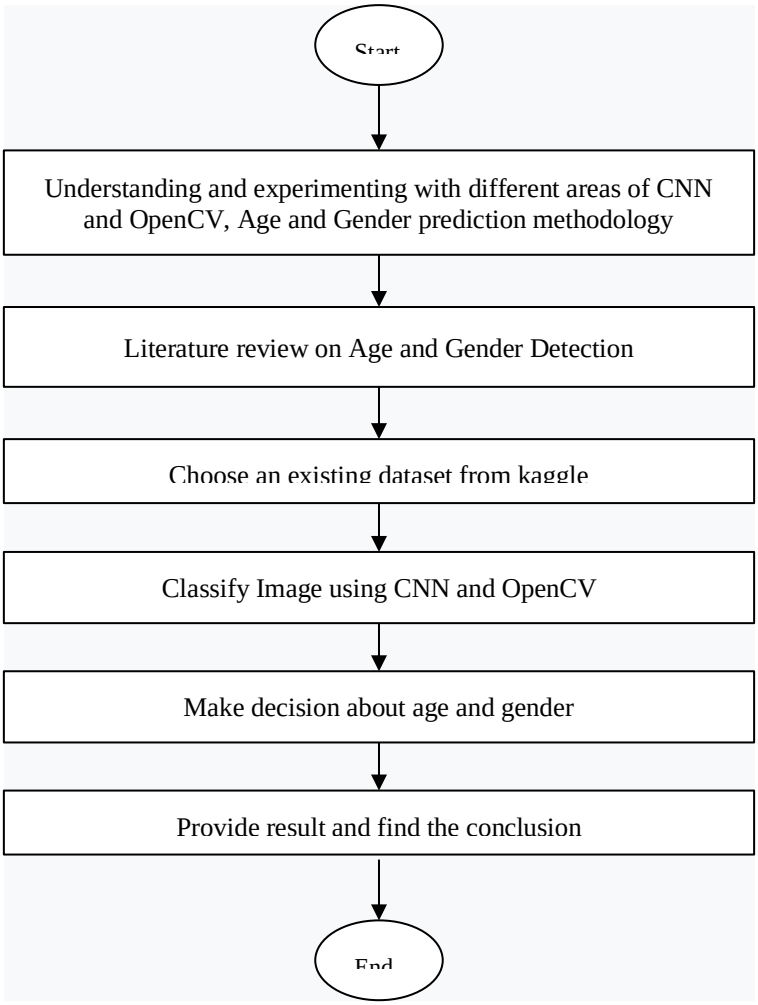

Figure 1.1 Research Flow

1.7 Significance of the research

We observe that most of the Age and Gender Detection datasets on the web are built on random pictures of humans. However, to analyse Age and Gender, we needed a dataset with a huge amount of pictures. Therefore we use CNN approaches to identify Age and Gender. Nowadays, automated markets/shopping malls require Age and Gender data. Age and Gender Detection can help to detect age and gender of people. It can help to detect age and gender in any field of our daily life.

1.8 Research Contribution

The overall contribution of the research work are :

- The project requires study of research on CNN and database learning.
- Data collection and information.
- Engineering design and development.
- Knowledge of software engineering and image processing.
- In this project where high definition images will be used, but the system will be able to predict even low quality images.

1.9 Thesis Organization

The thesis work is rganized as follows.

Chapter 2 highlights the background and literature review on the field of Age and Gender Detection. Chapter 3 contains Age and Gender Detection proposed architecture and a detailed walk-through of the overall procedures. Chapter 4 includes the details of the tests and evaluations performed to evaluate our proposed architecture. Chapter 5 explains the Standards, Impacts, Ethics, Challenges, the Constraints, Timeline, and Gantt Chart. Finally, Chapter 6 contains the overall conclusion of our thesis work.

1.10 Summary

This chapter includes a broad overview of the problem that we aimed explicitly at our research work's objectives, the background, and the research work's motivation. This chapter also illustrates the overall steps on which we carried out our research work.

Chapter 02

Literature Review

2.1 Introduction

This section succinctly reviews the associated works of age and gender estimation. Although a significant amount of literature is already available related to these topics, we try to provide a superficial outline in what way these tasks were approached earlier by other researchers.

2.2 Literature Review

Age estimation is a long-studied research topic among computer vision researchers. Most of the researchers considered human age estimation as either a classification or regression problem. In the case of age classification, age is coupled with a specific range or age group. On the other hand, age regression is a single value estimated for a person. However, it is very challenging to estimate an exact age due to diversity in the aging process across different ages. Furthermore, for accurate age estimation, the model needs a huge amount of correctly labeled face data. Many of the early age estimation methods used hand-crafted facial features in the constrained imaging conditions. A survey of such methods was reported in and a recent survey of age estimation including all approaches of the last decades can be found in . A method that extracts the geometric features from the face and calculates ratios among the facial features to estimate the age is presented in [28]. Initially, the face wrinkles are detected and localized, then the size and distances of the facial features are measured and finally the face is classified into different age categories. A similar approach as presented in [28] is proposed in [29] by modeling growth-related shape variations observed in human faces considering anthropometric evidence. This work was limited to a certain age. The above-mentioned methods are unsuitable for images in-the-wild due to the necessity of accurate localization of facial features. A couple of subspace methods were introduced, where aging features were extracted from an aging pattern representative subspace and a robust regression was used to predict the face ages. Although the aforementioned methods achieved excellent performances compared to previous cutting-edge methods, some limitations are exposed by these algorithms. Their system worked well only with frontal and properly aligned images. The algorithms proposed by these researchers are not well suited for practical applications where the input images might be collected in an unconstrained environment.

There are some methods where face images are represented using spatially localized facial patches. In , the patch distribution was represented by exploiting probabilistic Gaussian mixture models . A robust descriptor was used in place of pixel patches in . Later on, the Hidden-Markov-Model was introduced instead of the Gaussian mixture model (GMM) for representation of face patch distribution . A significant amount of research employed different robust image descriptors as an

alternative to the local image intensity patch for the age estimation task. Gao et al. used Gabor filters along with a Fuzzy-LDA classifier where one face belongs to multiple age classes. Similarly, Biologically-Inspired-Features in , and Local Binary Patterns (LBP) were presented in . The age estimation problem was considered as a regression problem by Fu et al. or as a classification problem using a quadratic function, shortest distance, and neural network-based classifiers in . The popular regression techniques reported by the researchers are Support Vector Regression (SVR) , Partial Least Squares (PLS) , Canonical Correlation Analysis (CCA) . Accordingly, Nearest Neighbor (NN) and Support Vector Machines (SVM) are the most used classification techniques. Next, we choose a couple of real age estimation methods to describe those that are most related to our suggested method. Guo et al. presented a learning scheme to draw aging features named manifold learning and utilized SVRs with local adjustment for age prediction, Han et al. extracted features using boosting algorithms and formed a hierarchical approach for classification between age group and regression inside a group (DIF). Geng et al. introduced an aging pattern subspace (AGES) wherefrom features were extracted and performed regression for age estimation. Zhang and Yeung handled age estimation as a multi-task problem based on the warped Gaussian process (MTWGP) where common features were shared among the tasks. Chen et al. Introduced a mapping between the cumulative attribute space and low-level sparse features for age regression. Chang et al. Formed the age labels into binary groups that formed subproblems and imposed a cost on each subproblem. Thus, they ranked the ordinal hyperplanes based on classification cost for age estimation, while Guo and Mu used a canonical correlation analysis and partial least squares founded methods to perform feature projection and estimate human traits jointly. Recently, the biologically inspired CNN models were successfully deployed for the age estimation task. Yi et al. [51] deployed a multiscale CNN. Wang et al. [52] used features from the intermediate layer of CNN rather than top layer features and performed manifold learning. Rothe et al. Incorporated a deep CNN for extracting features and real age regression as an estimate using SVR. In, they used a deeply learned CNN model from large in-the-wild image data and performed age regression through classification. A hybrid system was introduced, where the CNNs were used for face feature extraction and an extreme learning machine (ELM) for the classification task. A lightweight CNN network with mixed attention mechanism for low end devices was proposed in , where the output layer was fused by classification and regression approach. Another multi-task learning approach merging classification and regression concepts to fit the age regression model with heterogeneous data with the help of two different techniques for partitioning data towards classification was proposed in . To resolve the problem of data disparity and ensure the generality of the model, a very recent method is proposed by Kim et al. Where a cycle generative adversarial network-based race and age image transformation method is used to generate sufficient data for each distribution. All of the aforementioned CNN-based systems are evaluated on the basis of the common dataset Morph for age estimation. To the best of our knowledge, demonstrate state-of-the-art results. A lot of progress has been achieved in the gender estimation topic, but it is still a challenging problem in the real-world environment. The literature about gender estimation comes under the umbrella of the authors of . Here, we will discuss some of those methods where the well-known classifiers are used for gender estimation. As one of the very early methods, the authors used a fully connected two-layer neural network that learned from a limited number of near-frontal face images for gender classification. SVM classifiers were directly applied to image intensities. Similarly, AdaBoost was introduced instead of the SVM classifier by keeping the same working pipeline . Later on, a viewpoint-invariant model for age and gender estimation was suggested by Toews and Arbel which is robust to local scale rotations. A combination of human knowledge and a gait information-based gender

classification system was provided by Yu et al. An unconstrained face image benchmark for gender classification along with a high classification accuracy was presented in. Khan et al. Formed a semantic pyramid by extracting features from the full and upper body together with face regions from the image to recognize the gender and action. This method does not depend on the annotation value of a person's face and upper body to extract semantic features. In , the authors proposed a model where the first name of a person is used as a special feature and associates a name with facial appearance to recognize the gender of that person. At the same time, the authors showed that their method achieved higher accuracy in the task of gender recognition and demonstrated the potential in the use of face verification. In recent times, a generic framework for age and gender estimation was proposed in, where a hierarchical estimator was modeled based on extracted biologically inspired features. Besides, this method was formed to detect low-quality images due to a poor image background. The above efforts regarding gender estimation contributed a lot in this research area. However, the lion's share of these methods is only suitable for the applications with constraint images or have higher computational costs. Recently, a deep CNN-based approach was presented in. It was pretrained with a huge unconstrained dataset and then fine-tuned on two other datasets to achieve a very good accuracy in the gender estimation task. Although, their method states a high accuracy, a lot of pretraining is required prior to evaluating the system. In our paper, we propose a system that will work comparatively well with unconstrained imaging conditions.

2.3 Problem Analysis

We find CNN, SVM, KNN algorithms from the research paper where they did Age and Gender Classification. The researcher gets better performance using Convolutional Neural Network (CNN) architecture that is based on ANN. ANN is very popular in classification. ANN techniques are more suitable than other techniques.

2.4 Summary

This chapter includes an overview of the problem that we specifically target, the objectives of our thesis work along the motivation of the output of the thesis work. This section also elaborates on the overall steps on which we carried out our thesis work.

Chapter 03

Proposed Model

3.1 Introduction

This section presents the proposed deeply learned classifiers for age group and gender classification of unfiltered real-life face images, an extension of our conference paper in .

The approach as presented in Algorithm 1 requires image preprocessing (face detection, landmark detection, and face alignment) stage that preprocess and prepare the face images before they are input into the proposed network. Therefore, our solution is divided into three major steps: image preprocessing, features learning, and classification itself.

Intelligent age and gender classifiers tackle the classification task under unfiltered real-world settings. Most of those face images are not aligned and non frontal and also with different degrees of variations in pose, appearance, lighting, and background conditions. Therefore, those in-the-wild face images need first to be detected, then aligned, and, finally, used as input for the classifiers.

3.2 Feasibility Analysis

1. Technical Feasibility –we use this technology in this program
    - Pycharm
    - Python IDE
    - PIP Installer

2. Economic Feasibility –

   There is no economic feasibility required because we find every software we use in this program as free. And also we have our required hardware available.

3.3 Requirement Analysis

The proposed model includes:

Very fast computing device.

Pycharm/Kaggle

Open-source library function to calculate performance.

3.4 Research Methodology

In this section, we discuss our proposed CNN model overall process of how it works. This section divided 4 steps other sub-steps that show the input and output process of our proposed model. Each step shows detailed works and proper explanation**.**

3.4.1 Architecture of our Model

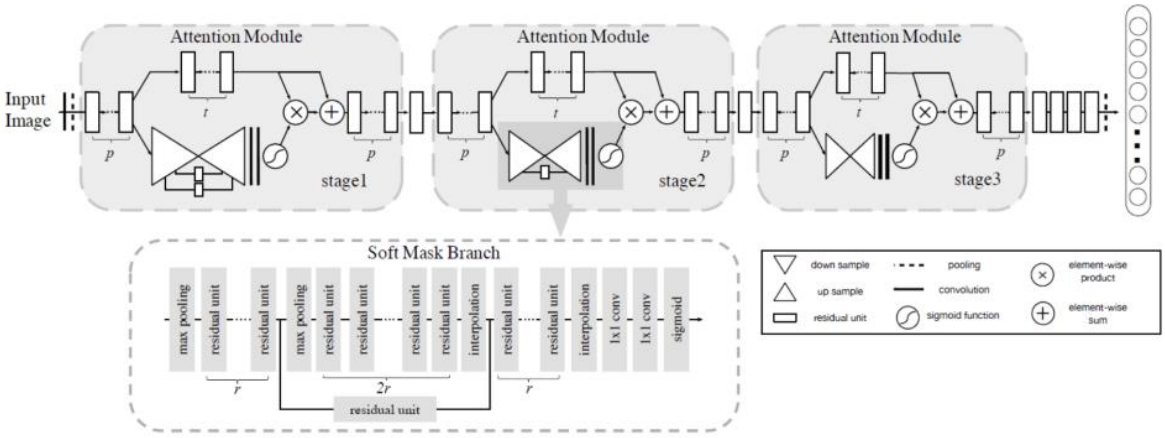

Figure 3.1: Architecture of our CNN Model

3.4.2 DFD in our model

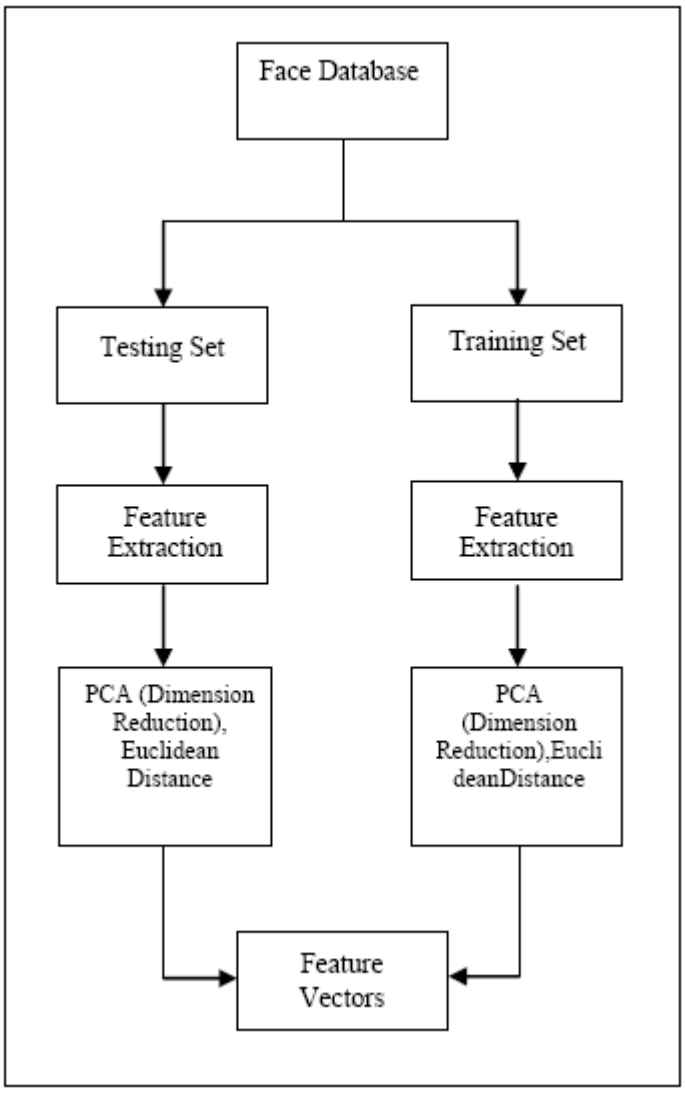

Figure 3.2: DFD in our Model

3.4.3 UML Design in our Model

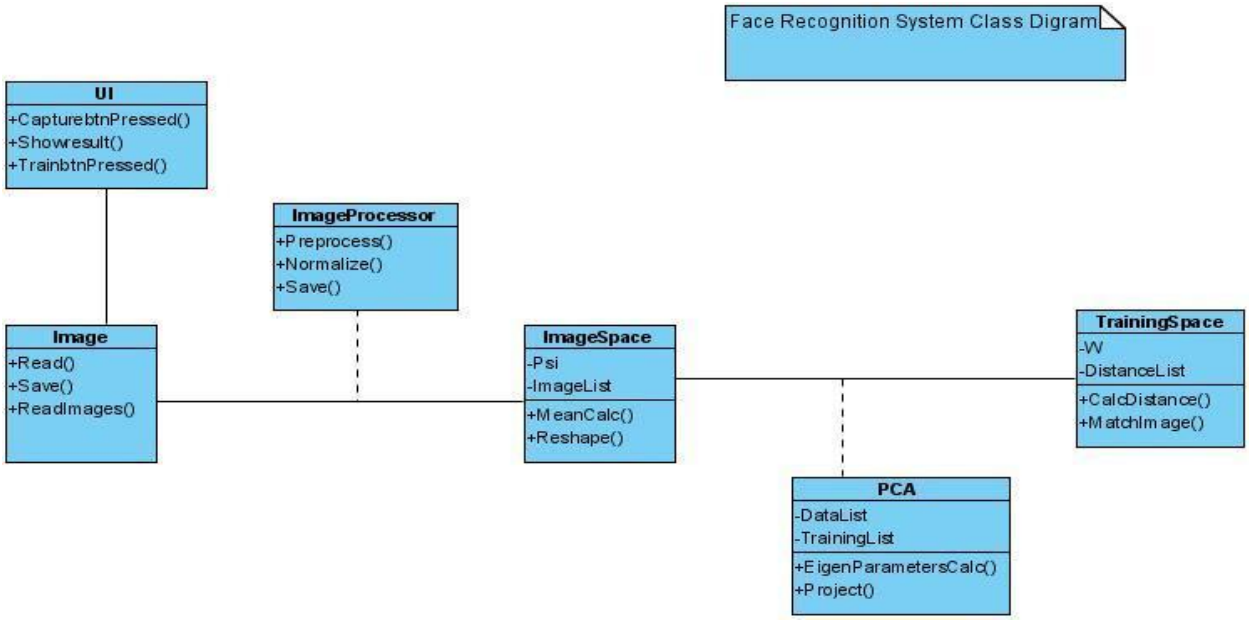

Figure 3.3: UML Design

3.5 Design, Implementation, and Simulation

Figure 3.1 explains the details. All steps of this model have implemented Python language and the details work throw platform is google collaborator. The Convolutional neural network models are implemented using python. Also, for additional calculation, and support, Scikit, and Numpy-pandas are used. The visual evaluation reports were generated using Matplotlib in python**.**

3.6 Summary

The proposed approach for estimating the real age and gender of a person demonstrates state-of-the-art results for the Morph dataset in anticipation of real age and the Adience dataset for gender estimation. We surpass the age estimation results marginally compared to the state-of-the-art DEX [24] method by lessening the data sparsity problem that exists in their approach. Our approach for the age estimation reduces the mean average error reported in the SOA method, i.e., the DEX fine-tuned for the IMDB-WIKI dataset and without fine-tuning by 0.26 years and 0.57 years, respectively. The system reveals that pre-training on a balanced dataset along with sufficient training data boosts the system performance reasonably. Our target application is a real-time system. Therefore, we incorporate the batch normalization layer before the non-linear RELU layer in the model to achieve a quicker convergence. In Table 3, it is empirically demonstrated that training the CNN model using a balanced dataset with batch normalization takes the lead in the result table. For the gender estimation, we used the ImageNet pre-trained model as the age estimation task without pre-training on the IMDB-WIKI dataset. We fine-tune and evaluate the model with the heterogeneous Audience data set and achieve SOA results. We achieve 5% more accuracy than the state-of-the-art approach [81]. In future research, we will try to enrich the balanced dataset because the amount of data in every class is not sufficient enough to train a very deep model and secure an enviable performance for a complex problem like real age estimation. We will try to devise an optimal network for the age and gender estimation from masked face images since smart store customers must wear a face mask amid the pandemic situation.

Chapter 04

Implementation and Testing

4.1 Introduction

This technique involves computation of a set of geometrical features such as nose width and length, mouth position and chin shape, etc. from the picture of the face we want to recognize. This set of features is then matched with the features of known individuals. A suitable metric such as Euclidean distance (finding the closest vector) can be used to find the closest match. Most pioneering work in face recognition was done using geometric features (Kanade, 1973), although Craw et al. (1987) did relatively recent work in this area. The advantage of using geometrical features as a basis for face recognition is that recognition is possible even at very low resolutions and with noisy images (images with many disorderly pixel intensities). Although the face cannot be viewed in detail its overall geometrical configuration can be extracted for face recognition. The technique's main disadvantage is that automated extraction of the facial geometrical features is very hard. Automated geometrical feature extraction based recognition is also very sensitive to the scaling and rotation of a face in the image plane (Brunelli and Poggio, 1993). This is apparent when we examine Kanade's(1973) results where he reported a recognition rate of between 45-75 % with a database of only 20 people. However if these features are extracted manually as in Goldstein et al. (1971), and Kaya and Kobayashi (1972) satisfactory results may be obtained.

4.2 System Setup

Automated age and gender estimation of the human face involves detecting, tracking and normalizing the face in an image sequence. The features extracted from the face image are used to estimate age and gender, and a face tracking method is employed to measure stay time and to count people. Figure 1 illustrates the system architecture used in this study.

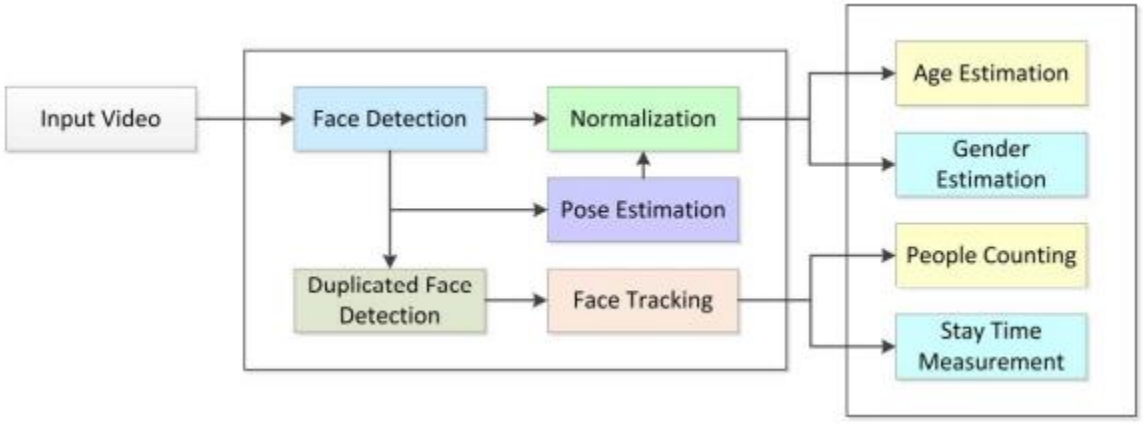

Figure 4.1: System architecture for age and gender estimation

4.3 Evaluation

We propose a model that uses CNN architecture to predict the age group and gender of human's faces from unfiltered real-world environments. The novel CNN approach addresses the age and gender labels as a set of discrete annotations and trains the classifiers that predict the human's age group and gender.

We design a quality and robust image preprocessing algorithm that prepares and preprocess the unfiltered images for the CNN model and this greatly has a very strong impact on the performance accuracy of our age and gender classifiers.

We demonstrate that pretraining on large-scale datasets allows an effective training of our age and gender CNN model which enable the classifiers to generalize on the test images and then avoid overfitting

Finally, OIU-Adience benchmark is used to evaluate the performance of our novel CNN model, and despite the very challenging nature of the images in the dataset, our approach produces significant improvements in age group and gender classification accuracy over the state-of-the-art methods; the result can satisfy the requirements of several real-world applications.

4.4 Results and Discussion

At the beginning of this section, we discuss the implementation details of the age and gender estimation task. Next, we introduce the datasets used for both tasks. Subsequently, we report both quantitative and qualitative results observed following the experiments. Finally, we discuss the results at the end of this section. We trained our deployed CNN model separately based on the age and gender classification task. We handle both tasks as a classification approach. The models are trained using the deep learning framework.In the case of age classification, we calculate the expected value from the Softmax probabilities belonging to output neurons, similar to what was considered in. We train the model with a balanced IMDB-WIKI dataset and introduce the batch normalization layer to reduce the training time. We consider every age as an individual class that ranges from 0 to 100. For all experiments regarding age estimation, the CNN is initialized with the weights trained on ImageNet . This pre-trained model is then further trained on the IMDB-WIKI image dataset for classification with 101 output neurons. Finally, the CNN is fine-tuned on the test dataset. For the gender classification, we report the gender class of the neuron carries the highest probability. We first deployed the pre-trained model trained on ImageNet. In the next step, we fine-tuned the pre-trained model with the real-world face image dataset added with two output neurons and the performance is reported from the test split of the Adience dataset. The training set consists of 80% of the images from the dataset and 20% is reserved for the testing. Further 90% of images from the training set are used for learning the weights and the rest of the images are used as validation sets during the training phase. Every experiment begins in conjunction with pre-trained ImageNet weights. When fine-tuning the pre-trained network with a smaller dataset, the learning rate of 0.001 remains fixed except for the last layer. The last layer weights are initialized randomly as the number of output neurons are changing. We used the Adam optimizer with the setting of momentum 0.9 and a weight decay rate of $5 \times 10{-}4$. We adjust the learning rate by a factor of 10 after every 30 epochs.

In this paper, we use four different datasets for real age and gender estimation. We first introduce the datasets with a description of their specifications. Figure 7 represents exemplar images for each dataset used for the age estimation experiment and sample images for gender prediction are presented in Figure 8. Table 4 shows the size of each dataset with its properties. For the age classification, we trained our model with the IMDBWIKI dataset and fine-tuned it with the

MORPH dataset. We evaluate our age estimation model with the MORPH dataset. In the gender estimation task, we train and evaluate our model with the dataset.Figure-Exemplar images with real (biological) age from each dataset used in the age estimation experiment.

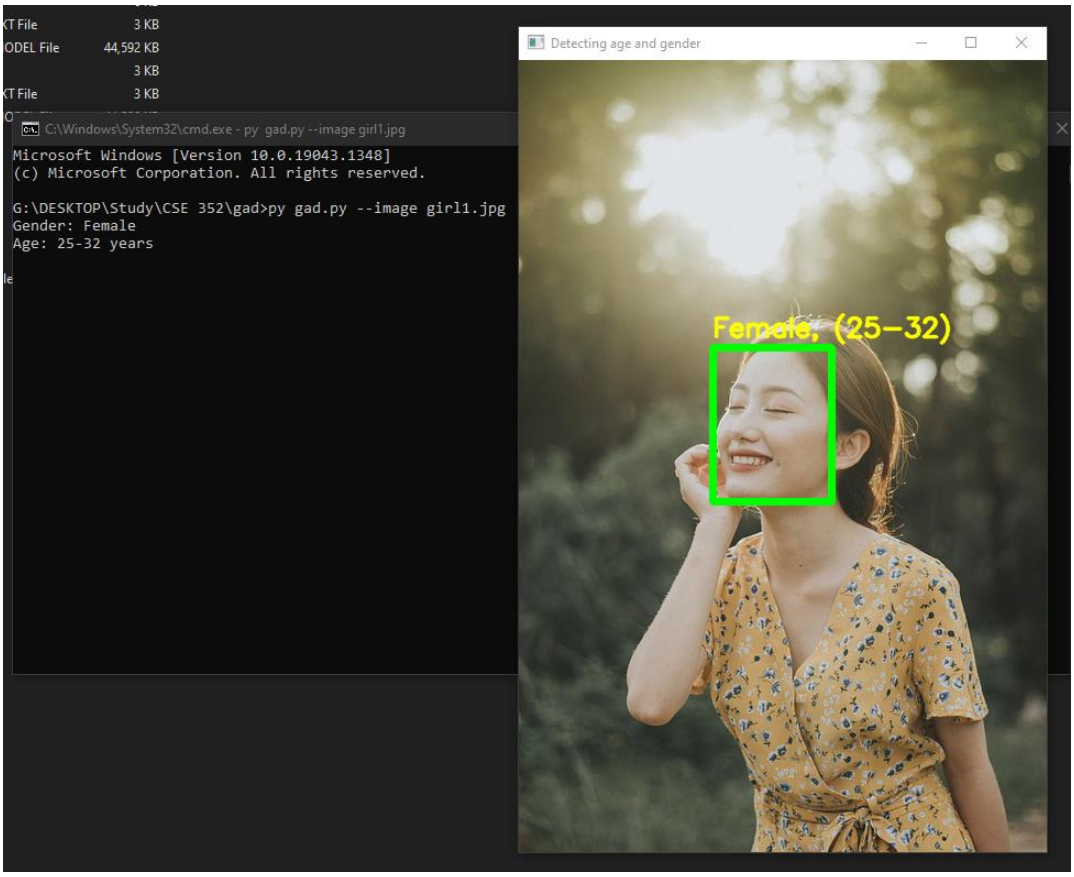

Figure 4.2: Result Example

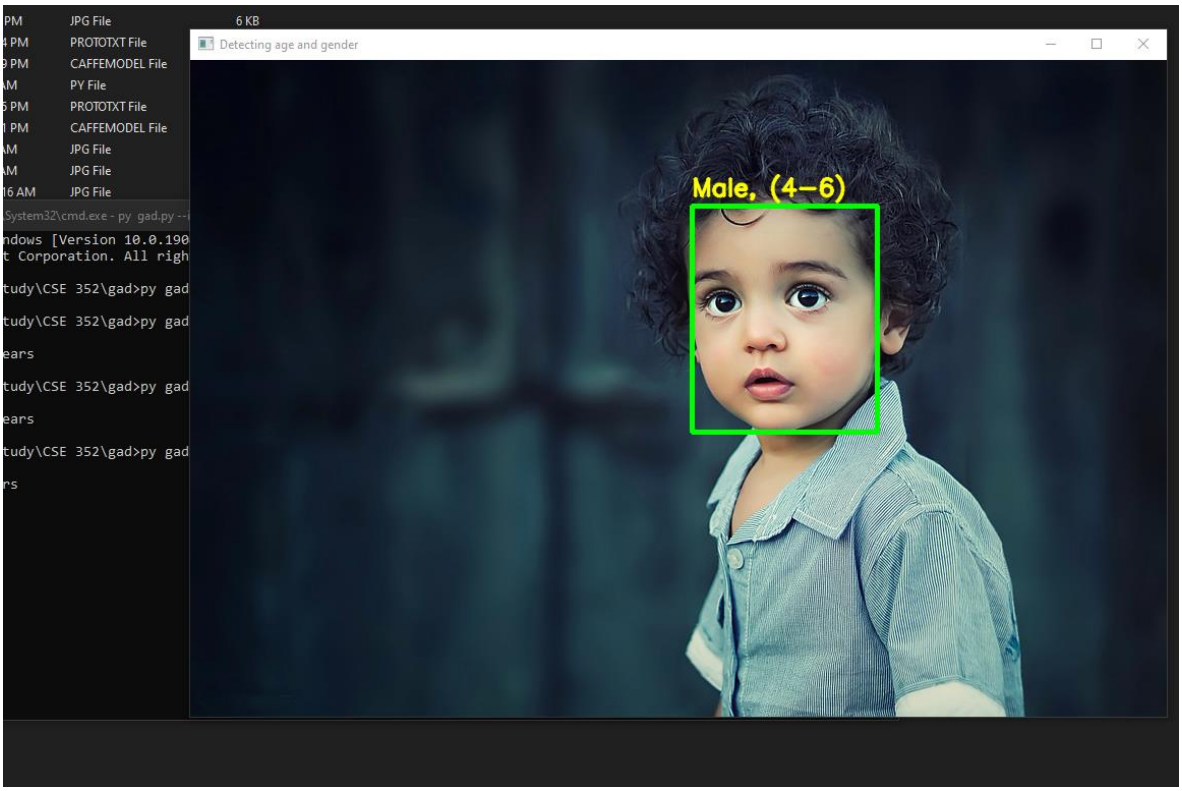

Figure 4.3: Result Example

4.5 Summary

The proposed approach for estimating the real age and gender of a person demonstrates state-of-the-art results for the Morph dataset in anticipation of real age and the dataset for gender estimation. We surpass the age estimation results marginally compared to the state-of-the-art DEX method by lessening the data sparsity problem that exists in their approach. Our approach for the age estimation reduces the mean average error reported in the SOA method, i.e., the DEX fine-tuned for the IMDB-WIKI dataset and without fine-tuning by 0.26 years and 0.57 years, respectively. The system reveals that pre-training on a balanced dataset along with sufficient training data boosts the system performance reasonably. Our target application is a real-time system. Therefore, we incorporate the batch normalization layer before the non-linear RELU layer in the model to achieve a quicker convergence. In Table 3, it is empirically demonstrated that training the CNN model using a balanced dataset with batch normalization takes the lead in the result table. For the gender estimation, we used the ImageNet pre-trained model as the age estimation task without pre-training on the IMDB-WIKI dataset. We fine-tune and evaluate the model with the heterogeneous dataset and achieve SOA results. We achieve 5% more accuracy than the state-of-the-art approach . In future research, we will try to enrich the balanced dataset because the amount of data in every class is not sufficient enough to train a very deep model and secure an enviable performance for a complex problem like real age estimation. We will try to devise an optimal network for the age and gender estimation from masked face images since smart store customers must wear a face mask amid the pandemic situation.

Chapter 05

Conclusion

5.1 Conclusions

In this paper, we propose an automatic age (biological) and gender estimation system for the promising smart store enterprise which is a modern venture in the retail industry. This automated system can extract the human demographics necessary to provide the customer a very good shopping experience that results in a boost of offline smart store sales. In addition, this enterprise solution can ease the shopping process and shorten the shopping time for the consumers. Although recent methods show their potential for the problem of age and gender estimation, the best works focused on constrained image benchmarks. As a result, these methods are not robust enough for the application involving real-world images. Most recently, some of the researchers learned to apply their model utilizing unconstrained image datasets, but these models are biased for the early and middle adulthood classes due to image sparsity problems in the dataset. In our paper, we resolve the data sparsity problem that exists in the state-of-the-art real age estimation approach DEX by constructing class-wise data parity and incorporate batch normalization concepts that jointly improve the previous results marginally. We follow the same setup for the gender estimation task and achieve substantially improved results over SOA methods regarding this task.

Author Contributions: Conceptualization, M.M.I. and J.-H.B.; methodology, M.M.I.; software, M.M.I.; validation, M.M.I. and J.-H.B.; formal analysis, M.M.I.; investigation, M.M.I; data curation, M.M.I.; writing—original draft preparation, M.M.I.; writing—review and editing, J.-H.B. and M.M.I.; visualization, M.M.I.; supervision, J.-H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the GRRC program of Gyeonggi province [GRRC Aviation 2017-B04, Development of Intelligent Interactive Media and Space Convergence Application System].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors have used publicly archived IMDB-WIKI and Adience dataset for the experiments. The IMDB-WIKI dataset is available in . The Audience dataset is available in [9].

Acknowledgments: We would like to acknowledge Korea Aerospace University with much appreciation for its ongoing support to our research. We are thankful to authors who make their datasets publicly available to pave the way to research in this area.

Conflicts of Interest: The authors declare no conflict of interest.

5.2 Future work

In our system we have used the modern web technologies to make our system fast, convenient and efficient for all of the personnel mentioned.Due to time and cost constraint it was not possible to fulfill all requirements and functionalities those were planned. But in future these planned functionalities and more improvement will be possible to pursue. The functionalities to be implemented areThus, there is a place for some future work such as:

● Different classifications models can be used at a time to improve the performance of the segmentation.

●To reduce the complexity of the algorithm,it's better to reduce the number of hypothesis to function the algorithm faster.

●To reduce the computation time, better filters are to be used to eliminate the unnecessary segmentation hypothesis.

●In future we want to implement distributed database.

●We will develop an android and iOS platform application.

●We will try to make easier and more reliable and handier user interface so that users can easily operate this application.

We also want to modify our application for the Future Work, we hope this work will help us in our future work. Our system is developed based on demand of user's satisfaction and facilities. In our system we have used the modern web technologies to make our system fast, convenient and efficient for all of the personnel mentioned. Due to time and cost constraint it was not possible to fulfill all requirements and functionalities those were planned.

Reference

1. Ricanek, K.; Tesafaye, T. Morph: A Longitudinal Image Database of Normal Adult Age-Progression. In Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 10–12 April 2006; pp. 341–345.

2. Geng, X.; Zhou, Z.-H.; Zhang, Y.; Li, G.; Dai, H. Learning from Facial Aging Patterns for Automatic Age Estimation. In Proceedings of the ACM International Conference on Multimedia, Santa Barbara, CA, USA, 23–27 October 2006; pp. 307–316.

3. Minear, M.; Park, D.C. A lifespan database of adult facial stimuli. Behav. Res. Methods Instrum. Comput. 2004, 36, 630–633.

4. Choi, S.E.; Lee, Y.J.; Lee, S.J.; Park, K.R.; Kim, J. Age Estimation Using a Hierarchical Classifier Based on Global and Local Facial Features. Pattern Recognit. 2011, 44, 1262–1281.

5. Song, Z.; Ni, B.; Guo, D.; Sim, T.; Yan, S. Learning Universal Multi-View Age Estimator Using Video Context. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 241–248.

6. A. Kumar and F. Shaik," Importance of Image Processing", 2016.

7. Chenjing Yan, Congyan Lang, Tao Wang, Xuetao Du, and Chen Zhang," Age Estimation Based on Convolutional Neural Network", 2014 Springer International Publishing Switzerland.

8. Hang Qi and Liqing Zhang," Age Classification System with ICA Based Local Facial Features", 2009 Springer-Verlag Berlin Heidelberg.

9. Eran Eidinger, Roee Enbar, and Tal Hassner," Age and Gender Estimation of Unfiltered Faces", 2014 IEEE.

10. Aditya K. Saxena, Shweta Sharma and Vijay K. Chaurasiya," Neural Network based Human Age-group Estimation in Curvelet Domain", 2015 The Authors Published by Elsevier.

11. J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao. Wld: "A robust local image descriptor", Trans.Pattern Anal. Mach. Intell.,

12. S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim. "Age estimation using a hierarchical classifier based on global and local facial features", Pattern Recognition, 44(6):1262–1281, 2011

13. T. F. Cootes, G. J. Edwards, and C. J. Taylor. "Active appearance models", In European Conf. Comput. Vision, pages 484–498. Springer, 1998.

14. C. Cortes and V. Vapnik. "Support-vector networks", Machinelearning, 20(3):273–297, 1995.

15. E. Eidinger, R. Enbar, and T. Hassner. “Age and gender estimation of unfiltered faces”, Trans. on Inform.Forensics andSecurity, 9(12), 2014.

16. Saragih, J.M.; Lucey, S.; Cohn, J.F. Deformable model fitting by regularized landmark mean-shift. Int. J. Comput. Vis. 2011, 91, 200–215.
Appl. Sci. 2021, 11, 4549 21 of 23

17. Xiong, X.; De la Torre, F. Supervised Descent Method and Its Applications to Face Alignment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 532–539.

18. Davies, G.; Ellis, H.; Shepherd, J. Perceiving and remembering faces. Am. J. Psychol. 1983, 96, 151–154.

19. Sinha, P.; Balas, B.; Ostrovsky, Y.; Russell, R. Face recognition by humans: Nineteen results all computer vision researchers should know about. Proc. IEEE 2006, 94, 1948–1962.

20. Gross, R.; Matthews, I.; Baker, S. Generic vs. person specific active appearance models. Image Vis. Comput. 2005, 23, 1080–1093.

Appendices

```
import cv2
import math
import argparse

def highlightFace(net, frame, conf_threshold=0.7):
    frameOpencvDnn=frame.copy()
    frameHeight=frameOpencvDnn.shape[0]
    frameWidth=frameOpencvDnn.shape[1]
    blob=cv2.dnn.blobFromImage(frameOpencvDnn, 1.0, (300, 300), [104, 117, 123], True,
False)

net.setInput(blob)
detections=net.forward()
faceBoxes=[]
for i in range(detections.shape[2]):
confidence=detections[0,0,i,2]
if confidence>conf_threshold:
x1=int(detections[0,0,i,3]*frameWidth)
y1=int(detections[0,0,i,4]*frameHeight)
x2=int(detections[0,0,i,5]*frameWidth)
y2=int(detections[0,0,i,6]*frameHeight)
faceBoxes.append([x1,y1,x2,y2])
cv2.rectangle(frameOpencvDnn, (x1,y1), (x2,y2), (0,255,0), int(round(frameHeight/150)), 8)
return frameOpencvDnn,faceBoxes


parser=argparse.ArgumentParser()

parser.add_argument('--image')

args=parser.parse_args()

faceProto="opencv_face_detector.pbtxt"
faceModel="opencv_face_detector_uint8.pb"
ageProto="age_deploy.prototxt"
ageModel="age_net.caffemodel"
genderProto="gender_deploy.prototxt"
genderModel="gender_net.caffemodel"

MODEL_MEAN_VALUES=(78.4263377603, 87.7689143744, 114.895847746)
ageList=['(0-2)', '(4-6)', '(8-12)', '(15-20)', '(25-32)', '(38-43)', '(48-53)', '(60-100)']
genderList=['Male','Female']
faceNet=cv2.dnn.readNet(faceModel,faceProto)
```

```
ageNet=cv2.dnn.readNet(ageModel,ageProto)
genderNet=cv2.dnn.readNet(genderModel,genderProto)

video=cv2.VideoCapture(args.image if args.image else 0)
padding=20
while cv2.waitKey(1)<0:
hasFrame,frame=video.read()
if not hasFrame:
cv2.waitKey()
break

resultImg,faceBoxes=highlightFace(faceNet,frame)
if not faceBoxes:
print("No face detected")

for faceBox in faceBoxes:
face=frame[max(0,faceBox[1]-padding):
min(faceBox[3]+padding,frame.shape[0]-1),max(0,faceBox[0]-padding)
:min(faceBox[2]+padding, frame.shape[1]-1)]

blob=cv2.dnn.blobFromImage(face, 1.0, (227,227), MODEL_MEAN_VALUES,
swapRB=False)
genderNet.setInput(blob)
genderPreds=genderNet.forward()
gender=genderList[genderPreds[0].argmax()]
print(f'Gender: {gender}')

ageNet.setInput(blob)
agePreds=ageNet.forward()
age=ageList[agePreds[0].argmax()]
print(f'Age: {age[1:-1]} years')

cv2.putText(resultImg, f'{gender}, {age}', (faceBox[0], faceBox[1]-10),
cv2.FONT_HERSHEY_SIMPLEX, 0.8, (0,255,255), 2, cv2.LINE_AA)
cv2.imshow("Detecting age and gender", resultImg)
```