Human Gender and Age Detection Through Image Processing Technique

A project Paper submitted in partial fulfilment of the requirement for the Degree of Bachelor of Science (Engg.) in Information and Communication Technology

Submitted By

ID : 12009022

ID : 12009031

Session : 2019-20

Supervised By

Amena Begum

Assistant Professor

Department of Information and Communication Technology

Comilla University

Submission: February,2025

Acknowledgement

It is a great pleasure for us to express unbound indebted gratitude to our honorable supervisor, **Amena Begum, Assistant Professor, Department of Information & Communication Technology, Comilla University** for her numerous helpful & constructive suggestions, scholastic guidance, constant inspiration, valuable advice, and kind co-operation for the successful completion of work "Human Gender and Age Detection Through Image Processing Technique". She has always been extremely generous with her time, knowledge, and ideas allowing us great freedom in this research. Her enthusiastic approach to research, endless excitement for this project, and effervescent personality have made this project all the more enjoyable, and we are greatly encouraged.

We are also grateful to all the other teachers in the department for their inspiration during this Study for giving us the proper guideline throughout the entire semester. Finally, we would like to convey our special thanks and gratitude to our beloved family members and friends for offering consistent support in many ways. They will always be remembered and deeply appreciated by us.

Abstract

Interest in automatically classifying people's age and gender from face photographs has grown since the introduction of social media. Thus, the age and gender categorization process is an essential step in many applications, including interest group targeting, aging analysis, face verification, and ad targeting. However, the majority of age and gender classification systems still have some issues when used in practical settings.

This study employs several convolutional neural networks (CNN) for the categorization of age and gender. The suggested approach consists of five stages: multiple CNN, face alignment, background reduction, face identification, and voting systems. In terms of structure and depth, the multiple CNN model has three distinct CNNs; the objective of this variation is to extract distinct characteristics for each. This is accomplished by using machine learning models and algorithms that have been trained on big datasets of photos of individuals of various ages.

Our project, " Gender and Age Detection Through Image Processing Technique," makes use of intelligent technologies to infer an individual's age from their photos. Our goal is to educate computers to identify age-related facial traits so they may be used for demographic analysis, content recommendation, and targeted advertising. Our initiative seeks to be accurate and useful in real-world scenarios by utilizing cutting-edge methodologies. It's similar to teaching computers to make age predictions based only on a person's picture. true-time age prediction is where the true magic occurs. Consider submitting a photo, and our algorithm will provide you with an age estimate right away. In addition to this wow element, there are useful uses.

Table of Contents

CHAPTER THREE – METHODOLOGY

CHAPTER FOUR – SYSTEM DESIGN

CHAPTER FIVE – IMPLEMENTATION

CHAPTER SIX – RESULT AND DISCUSSION

CHAPTER SEVEN – CONCLUSION

List of Figures

List of Abbreviations

- CNN Convolutional Neural Network
- CUDA Compute Unified Device Architecture
- DNN Deep Neural Network
- GPU Graphics Processing Unit
- HOG Histogram of Oriented Gradients
- k-NN k-Nearest Neighbors
- LBP Local Binary pattern
- ReLU Rectified Linear Unit
- SSD Single Shot MultiBox Detector
- SVG Scalable Vectors Graphics
- SVM Support Vector Machine
- YOLO You Only Look Once

CHAPTER 1

INTRODUCTION

1.1 General Introduction

The fast development of technology in the realms of computer vision and machine learning has profoundly affected many spheres, including security, healthcare, marketing, and human-computer interaction. Among these developments, image processing's vast range of practical implementations makes age and gender detection especially important. From digital photos or video feeds, this technology uses facial recognition and image analysis methods to ascertain a person's gender and age.

Mostly driven by image processing methods mixed with machine learning algorithms, age and gender detection systems to project demographic information, these algorithms examine facial traits including eye shape, nose, jawline, skin texture, and other biometric indications. Deep learning model especially CNN—have improved the accuracy and efficiency of such systems still further [1].

Different applications clearly show the value of age and gender identification. For security-oriented surveillance systems, for example, it helps locate people. In the retail sector, companies use demographic information to customize marketing plans and enhance client interactions. Furthermore, important in social media platforms for content personalizing and in healthcare for patient monitoring is age and gender estimate [2].

Age and gender identification present various difficulties even if their importance is rising: differences in lighting conditions, facial expressions, occlusions, and aging effects over time. Dealing with these difficulties calls for both strong algorithms and varied datasets to guarantee dependability and great precision.

1.2 Significance of Gender and Age Detection

Gender and age detection have transforming effects in many sectors, from improving security to customizing user experiences to optimizing corporate objectives. For the following reasons this technology is absolutely essential. Modern security systems are built with age and gender detection systems right in mind. They help law enforcement authorities detect and identify people in public areas, therefore increasing reaction times during investigations, and crowd surveillance quality. In the retail and advertising sectors, demographic insights obtained from age and gender detection help companies to customize marketing campaigns, suggest products, and enhance customer engagement strategies, so improving sales and customer satisfaction. Particularly for patient monitoring, age-specific diagnoses, and treatment regimens, age estimate is absolutely vital in healthcare. Gender detection guarantees proper medical treatment and helps one to grasp gender-related health trends.

Social media platforms use age and gender detection to suggest material, increase user interaction, and guarantee compliance with age-restricted content rules, thereby improving the whole user experience by means of content customization. Age and gender detection in human-computer interaction (HCI) enhances application adaptability, hence enabling more responsive and intuitive user interfaces grounded on demographic data. Demographic detection systems are used by government and public organizations for census data collecting, urban planning, and resource allocation so guaranteeing improved public service delivery.
All things considered, the combination of gender and age recognition systems helps to make technology more user-centric, safe, and responsive, thereby promoting developments in many fields.

1.3 Aim & Objectives

This work aims to present a thorough description of the design, development, implementation, and assessment of a Gender and Age Detection System employing cutting-edge image processing and machine learning methods. The paper focusses on how different algorithms and models SSD for face detection, CNN for feature extraction, and DNN for classification are combined to produce accurate and dependable gender and age prediction. The paper also seeks to underline the performance measures of the system, its efficiency, and possible uses in practical settings. This work offers to:

- Clarifying the basic ideas underpinning face detection, feature extraction, and classification methods will help one to grasp how these methods support gender and age detection.
- Detailing how various components including SSD, CNN, and DNN cooperate to analyze input photos and get correct classification results, therefore describing the system architecture and suggested procedure.
- To detail the development and implementation process including the tools and technologies utilized, like Python, Open-CV, and libraries like argprse and os, together with the dataset employed (Adience Benchmark for Gender and Age Classification).
- Analyzing important criteria like accuracy, precision, recall, and F1-score can help one assess the system's dependability in several contexts and datasets.
- To pinpoint and talk about the shortcomings of the present system, including issues with different lighting situations, face occlusions, and dataset imbalances that might compromise the performance of the model.
- Suggesting means to increase the accuracy, resilience, and scalability of the system for more general uses in sectors such surveillance, security, human-computer interaction, and targeted marketing, helps one to propose future work and enhancements.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Because it can be applied in so many fields, including security monitoring, biometric identification, targeted advertising, and human-computer interaction, image processing for gender and age has grown to be a major area of research. These types of technologies mostly aim to rapidly view digital images of people's faces and deduce their gender and age. Different face expressions, lighting, occlusions, the effects of ageing, and varying demographics are some of the challenges this work presents.

Historically, hand-made features such SVM, HOG, LBP were employed for classification. These techniques performed effectively in controlled environments but struggled with factors like changing illumination, face expressions, and occlusions in actual life. Thanks to developments in deep learning and computer vision, gender and age detection systems have improved greatly throughout the past few years. For feature extraction and classification, CNNs come in really handy. From raw picture data, they can learn hierarchical representations straight forwardly. CNNs automatically identify intricate patterns, so models are more accurate and dependable than prior machine learning techniques depending on hand-made features.

Including the DNN functionality into Open-CV has also helped to simplify using deep learning models in real-time applications. This program supports several pre-trained models, which facilitates rapid development and enhancement of age and gender identification systems. The SSD approach performs quite well as an object and face detector as it can discover objects rapidly and precisely in a single forward run of the network. Some of the most significant research and techniques applied in determining someone's gender and age are discussed in this article. It emphasizes how these techniques have evolved over time from conventional image processing techniques to more contemporary deep learning approaches. The Open-CV, DNN module for fast model deployment, CNNs' responsibilities for learning features, and SSD for real-time facial detection are discussed in this paper. It provides a complete backdrop for appreciating the work of the project and fresh concepts

2.2 Existing Methods and Limitations

From conventional machine learning-based methods to more sophisticated deep learning techniques, gender and age recognition using face photos has seen tremendous change over time. The majority of early techniques employed hand-crafted features like Haar-like features, LBP, and HOG. These features were then identified using algorithms like k-NN and SVM. [3]. CNNs emerged as the primary method for classifying age and gender with the advent of deep learning. CNNs are excellent in automatically extracting features from data,
eliminating the need for human feature extraction. Their ability to automatically recognize intricate patterns significantly increases accuracy in a variety of unrestricted settings. However, in order to train well, CNNs need a lot of labelled data, which can be resource-intensive.

Deep network training is also computationally costly, requiring a lot of time for model optimization and high-performance hardware. These difficulties restrict its use in contexts with limited resources and real-time systems [4][5].

CNNs are also susceptible to biases in the training data. The model's performance may suffer if the dataset does not fairly reflect a variety of demographic groupings (such as various ages, genders, or ethnicities), particularly for under-represented groups. Concerns regarding equity and model generalization to a larger population are brought up by this problem . Initially, techniques like Viola-Jones and Haar cascades were employed for face detection. Despite being fundamental, these methods have trouble processing in real time and maintaining accuracy in a variety of scenarios, including occlusions and odd head positions. By enabling real-time processing, even on video streams, contemporary methods like SSD and YOLO have greatly increased face detection speed and accuracy. These techniques still have drawbacks, though, especially when dealing with low-resolution photos or faces that are hidden by masks or spectacles [6]. There are various methods which are used.

2.2.1 Traditional Machine Learning Methods

Before the arrival of deep learning models, conventional machine learning techniques such SVM and k-NN have been extensively applied for age and gender categorization chores. Unlike contemporary deep learning-based techniques, these methods depend on hand-crafted characteristics retrieved from input photos, which distinguishes them Designed to determine the ideal hyperplane separating data into several groups, SVM are supervised learning systems. In the framework of gender and age classification, SVM aims to divide male and female faces, or pictures corresponding to various age groups, depending on attributes derived from the facial images. [7]

Limitation: Manual feature extraction is required in these models, which can be error-prone and might not capture the intricate interactions in face characteristics. These techniques do not scale well with the rising complexity of contemporary datasets, so they are not appropriate for handling real-time applications or big-scale datasets. Variations in posture, illumination, and facial expressions cause generalization on real-world photographs to be frequently lacking in these techniques.

2.2.2 Convolutional Neural Networks

CNNs have transformed computer vision tasks including gender and age categorization. Instead of laborious feature extraction, CNNs learn hierarchical features from raw picture data. CNNs use convolutional, pooling, and fully connected layers to extract increasingly complex characteristics from the input picture. CNNs use convolutional layers to identify edges, textures, and corners by filtering the input picture. These filters build feature maps of learnt patterns by sliding over the picture in tiny chunks. Pooling layers reduce feature map dimensionality and preserve the most relevant information after convolutional layers. The last fully connected layers classify the retrieved characteristics and output the projected age or gender. [8]

Limitations: CNNs need on a lot of labelled data for training, hence in circumstances when annotated data is rare or costly to collect, this might be a major restriction. CNNs may over-fit the training data without appropriate regularizing or augmentation, therefore producing inadequate generalization to unseen pictures. Especially in safety-critical applications, CNNs' black-box character makes it challenging to comprehend the decision-making process behind predictions.

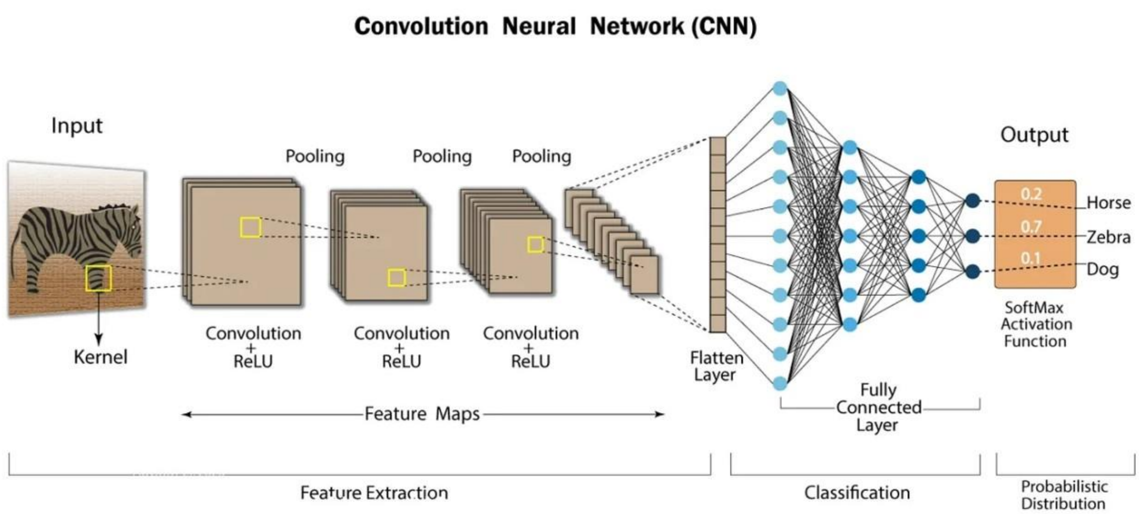

Figure 2.1 - Convolution Neural Network(CNN) Structure

2.2.3 Residual Networks

Designed to solve the difficulties of training extremely deep neural networks, ResNet Residual Network - is a deep learning architecture. Conventional deep neural networks often suffer from the vanishing gradient issue, in which case the network finds it difficult to train efficiently when the gradients get very tiny during backwards propagation. ResNet addresses this problem by adding residual connections, which let the network skip certain levels and straight forwardly transport data from one layer to the next. These residual connections provide "shortcuts" allowing the network to store more information, hence supporting the training of deeper networks without compromising performance.

Every block in ResNet consists of two or more convolutional layers with a skip connection avoiding these layers therefore adding the input to the convolutional output. This residual learning permits the model to concentrate on learning the difference—that is, residual—between the input and the output instead of learning the whole transformation, therefore enhancing the training of deeper models. Unlike conventional CNNs, the design lets one create networks with hundreds or even thousands of layers, hence enabling considerably deeper models. [9]

Limitations: ResNet designs may be computationally costly despite their great performance, which makes deployment challenging in settings with limited resources. Training extremely deep networks like ResNet calls for large computing resources, which may result in extended training periods and need specialized gear like GPUs or TPUs.

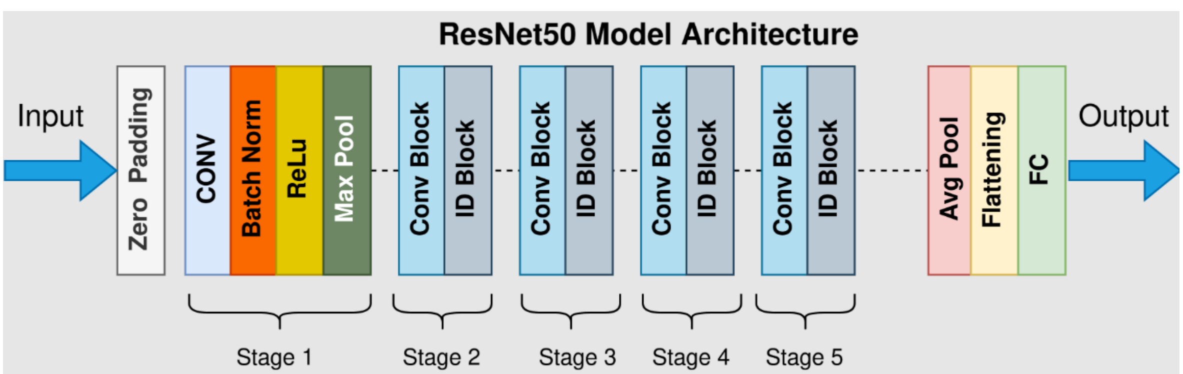

Figure 2.2 - ResNet 50 Model Architecture

2.2.4 Multi-Task Learning

Under the multi-task learning (MTL) machine learning paradigm, a model is taught to concurrently complete many related tasks while sharing knowledge across them to raise performance. Using shared characteristics to improve learning, MTL lets a model forecast both age and gender from the same set of data in the framework of age and gender categorization. Learning many tasks concurrently helps the model to generalize as it catches more varied patterns relevant for all jobs. Encouragement of the model to utilize common characteristics across many but related tasks has been proven to enhance performance in several applications, including image recognition. When activities are not properly connected, MTL may experience task interference—that is, when learning one activity reduces the performance of another. Notwithstanding this, MTL is a strong method particularly in cases of limited data as it lets the model profit from the common structure across jobs. [10]

Limitations: Training tasks together may lead to negative transfer, wherein the performance of one task influences the performance of another when tasks are not sufficiently connected. Particularly when the tasks are somewhat varied, the multi-task model is more complicated and may need more data to adequately train all tasks.

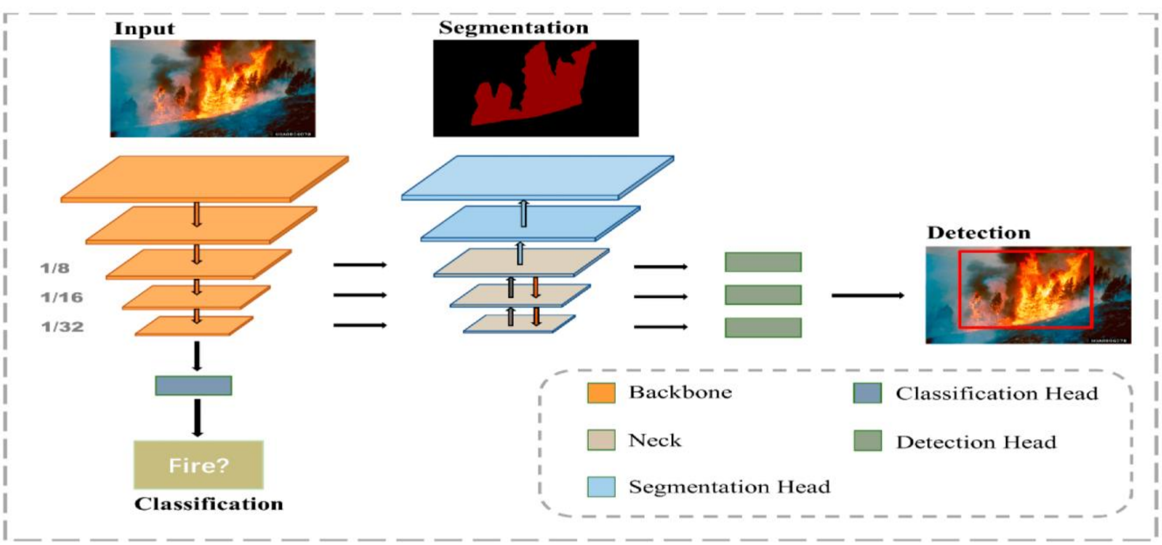

Figure 2.3 : Multitask Learning Model

2.2.5 Generative Adversarial Networks

Two neural networks a generator and a discrimination form the class of deep learning models known as generative adversarial networks (GANs). Whereas the discriminator determines if the data is real from the dataset or fake generated by the generator he generator generates synthetic data such as photographs. Together in a competitive process, these two networks are trained; the generator seeks to generate more realistic data to trick the discriminator, while the discriminator aims to improve at differentiating genuine from false data. This adversarial approach produces synthetic data of excellent quality that very nearly reflects actual data. Image production, data enrichment, and other model performance enhancement including those for age and gender classification have only been a few of
the uses for GANs. GANs may be computationally costly and challenging to train notwithstanding their success. [11]

Limitations: GANs are famously difficult to train and may experience mode collapse, in which case the generator fails to generate varied pictures. The GAN training procedure is computationally costly and usually calls for large resources and time. Synthetic data's quality may not always coincide with real-world data, particularly in dynamically fluctuating situations like ageing development.

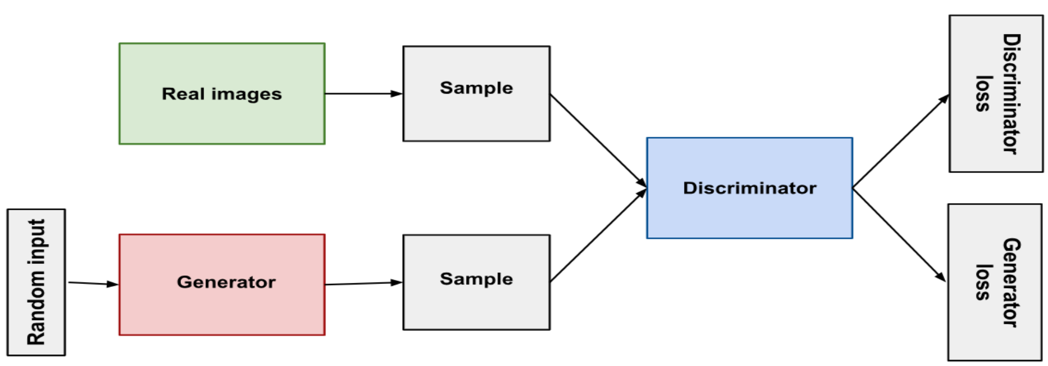

Figure 2.4 : Generative Adversarial Networks (GANs) Model

2.2.6 MobileNet:

Lightweight deep learning architecture meant for mobile and embedded devices is MobileNet. It breaks down the typical convolution process into two smaller operations: depthwise convolution and pointwise convolution using depthwise separable convictions. By lowering the computational complexity and the number of parameters, MobileNet is very efficient even if it still maintains competitive performance in jobs like picture categorization. Its efficiency lets it function on devices with limited processing capacity like cellphones without compromising too much accuracy. MobileNet has been extensively embraced for real-time applications where speed and accuracy are absolutely vital. [12]

Limitations: Although MobileNet is economical, in terms of accuracy especially its performance may not always equal more complicated designs like ResNet or VGGNet. MobileNet's lowered accuracy may not be enough for strong predictions in really demanding uses like real-time face recognition in congested surroundings.

2.2.7 Xception

Inspired on depthwise separable convolutions, like MobileNet, Xception is a deep learning architecture with a more intricate design. It maintains efficiency by substituting depthwise separable convolutions throughout the whole network for the conventional convolution operations, hence enabling learning of more intricate representations. Because Xception can more successfully capture complex patterns and features than conventional CNNs, it is especially well-known for its performance in picture classification challenges. Xception is appropriate for jobs requiring thorough feature extraction from big-scale datasets albeit more computationally demanding than MobileNet as it offers a solid mix of accuracy and efficiency. [13]

Limitations: Despite having fewer parameters than traditional CNNs, Xception still requires considerable computational resources, making it less suitable for real-time applications on low powered devices. The architecture is complex, and fine-tuning it for specific applications such as age and gender detection requires significant effort and expertise.

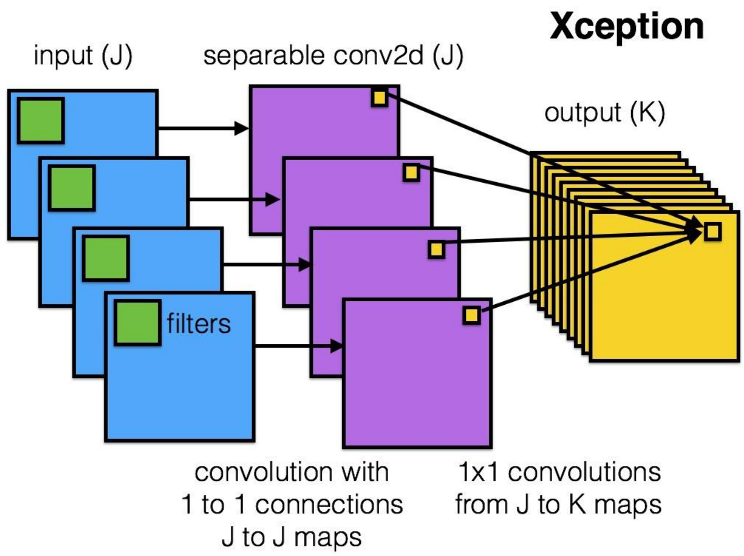

Figure 2.5 : Xception Architectural Design

2.3 Integration of CNN, SSD, DNN

With the use of Open-CV, DNNs, CNNs, and SSD, gender and age detection using face photographs has advanced significantly. For real-time applications, Open-CV, an open-source computer vision toolkit, is essential for face identification, preprocessing, and integration with machine learning models. The integration of CNNs for age and gender categorization has been the subject of several research. CNNs remove the need for manual feature extraction by automatically learning the spatial hierarchies of features from raw pictures. For instance, by using big datasets for training, Zhang et al.'s deep CNN architecture for joint age and gender prediction obtained great accuracy [6].

Similar to this, Simonyan and Zisserman demonstrated the efficacy of very deep CNNs in identifying minor traits that are crucial for classifying people according to their age and gender by using them for face recognition tasks. Pre-trained architectures like VGGNet and ResNet, which were optimized for certain applications, further improved these models and greatly increased classification accuracy. Because of its effectiveness and quickness, the SSD framework which was first presented by Liu et al has grown in popularity as a face identification method for photos. SSD is perfect for real-time processing in surveillance and monitoring systems since it can identify several objects and faces at once with no computing expense. Several gender and age detection systems have effectively used the combination of CNNs for classification and SSDs for face detection.

For example, Khan et al. achieved high accuracy in real-time performance by using SSD for quick face identification and a CNN model for gender and age prediction [7]. For real-world gender and age recognition applications, Open-CV's assistance in managing picture preprocessing, face detection, and integration with DNNs and CNNs has shown to be a useful and adaptable solution.

2.4 Discussion

Since deep learning was introduced, gender and age detection has changed dramatically, especially with the use of CNNs. Early approaches relied on conventional machine learning algorithms using handmade features, notably LBP and Haar-like features, which performed well in controlled settings but poorly in situations with changes in illumination, position, and emotions. Additionally, all approaches needed manual feature extraction, which limited their scalability and flexibility in practical settings.

By automating feature extraction, CNNs transformed the field and made it possible for models to directly learn intricate patterns from data. Accuracy increased as a result, especially for age and gender categorization under various circumstances. CNNs do, however, provide a unique set of difficulties. Large, labelled datasets are necessary for training these models, but they are sometimes hard to come by, particularly for varied populations. Furthermore, deep models can be prohibitively expensive to train computationally, which makes them unsuitable for real-time applications or devices with limited resources. Model bias is another problem, whereby deep learning models that were trained on skewed datasets could perform worse for particular demographic groups. Fairness and inclusion are called into question since the model's predictions can give preference to under-represented groups. Techniques like YOLO and SSD have increased real-time accuracy in face identification, making them appropriate for uses like video monitoring. They still have difficulties, though, especially when it comes to identifying identities in low-resolution photos or when faces are partially hidden. Accuracy and speed are still traded off in real time processing, with good performance frequently necessitating large computational resources. Overall, even though there has been progress, issues with data, computation, bias, and real-time performance still exist, necessitating more study and improvement to increase the efficiency and equity of age and gender identification systems

CHAPTER 3

METHODOLOGY

3.1 Introduction System Analysis:

The process used to create the Gender and Age Detection using Image Processing system, which uses cutting-edge machine learning methods to categorise people's age and gender from pictures of their faces. The system uses a variety of sophisticated algorithms, including CNN, DNN, and SSD for face identification, to automatically determine an individual's gender and age from an image. During the System Analysis phase, the project's requirements and limitations are assessed, and the right tools and techniques are chosen to satisfy those demands. This stage guarantees that the built system is accurate and computationally efficient, and that it can handle real-time face detection and age-gender categorization. Because deep learning models, particularly CNNs and DNNs, consume a lot of resources, the system is made to balance performance and resource use as best it can. While inference is optimized for real-time performance on modest hardware configurations, model training is carried out on high-performance platforms. The model also attempts to handle issues like position variation, illumination, and occlusions in real-world situations.

3.2 Face Detection

Face detection is a necessary phase in the gender and age identification system as it forms the basis of exact categorization. This work uses the well-known SSD technology, which is highly successful in real-time object recognition applications. While concurrently building many bounding boxes around objects in an image, SSD forecasts the class probabilities of them. It divides the input picture into a grid and use convolutional filters to find faces at different sizes and aspect ratios. SSD mathematically optimizes the multibox loss function, which combines bounding box regression's localization loss (Smooth L_1 Loss) with confidence loss (Softmax Loss). The loss function's recipe is:

$$\boldsymbol{L_{(x,c,l,g)}=N_1\ (L_{conf(x,c)}+\alpha L_{loc(x,l,g)})}$$

Where:

L_{conf} is the confidence loss,

L_{loc} is the localization loss,

N is the number of matched default boxes,

α is the weighting factor.

The SSD model enables fast, accurate face detection, making it ideal for real-time applications like video surveillance.

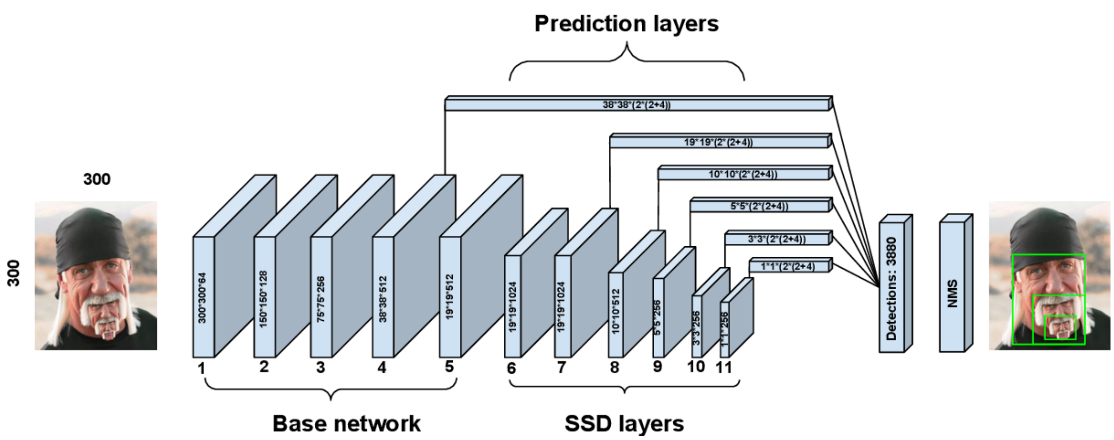

Figure 3.1 - Face Detection Using SSD

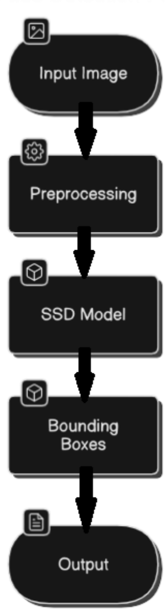

Figure 3.2 - SSD work procedure

3.3 Feature Extraction

In the gender and age identification system, feature extraction is a vital step where the objective is to recognize and depict important face traits that can differentiate between various age groups and sexes. This work uses CNNs for feature extraction as they can automatically learn spatial hierarchies of features from unprocessed pixel input. A CNN consists in several layers: convolutional layers, activation functions (ReLU), pooling layers,
and fully connected layers. Convolutional layers' learnable filters kernels applied to the input image provide feature maps highlighting important patterns like edges, textures, and facial landmarks (eyes, nose, mouth, etc.). Mathematical activity of convolution is described as follows. The method for pulling features can be written in math terms as a series of convolution operations:

$$\boldsymbol{S_{(i,j)}=\sum_m \sum_n I(m,n)K(i-m,j-n)}$$

Where:

• I represent the input image,

• K is the convolutional kernel,

• $S_{(i,j)}$ is the resulting feature map.

After convolution, the ReLU activation function introduces non-linearity, enabling the network to learn complex patterns. Pooling layers (e.g., max-pooling) reduce the spatial dimensions of feature maps, improving computational efficiency while preserving essential information.

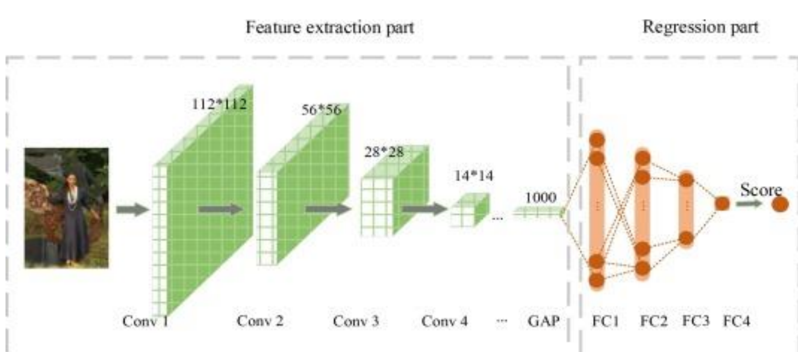

Figure 3.3 - Feature Extraction Using CCN

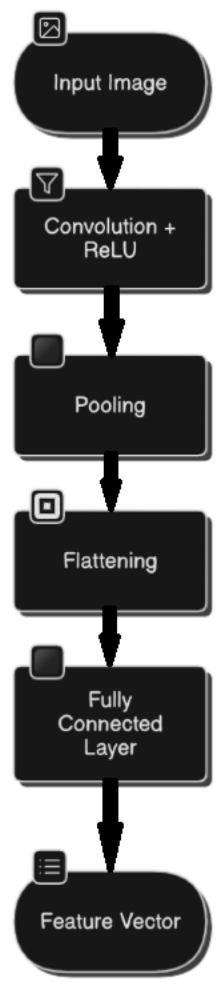

Figure 3.4 – CCN work procedure

3.4 Gender and Age Classification

Gender and age categorization is the last phase of the process in which the obtained facial characteristics are managed to project the gender and age group of the individual. We apply DNN in our work for classification as they can learn complex patterns and correlations from high-dimensional data. Comprising numerous layers an input layer, several hidden layers, and an output layer a DNN is through weighted connections, every layer consists of neurons linked with the one before and next. The DNN receives as input the feature vector derived from CNN-based feature extraction method. Accurate categorization depends on these fundamental facial characteristics shape, texture, and spatial relationships which our tools effectively capture. The following mathematical statement controls the fundamental running of a DNN:

$$\boldsymbol{a(1) = f(W(1)a(l-1) + b(1))}$$

Where:

• a(l) is the activation of the lth layer,

• W(l) represents the weights matrix,

• b(l) is the bias vector,

• f is the activation function (commonly ReLU for hidden layers and Softmax for the output layer).

By adding non-linearity via the ReLU (Rectified Linear Unit) activation function, the model may learn intricate decision limits. The output layer uses the softmax function to translate raw scores into probabilities, therefore enabling the model to make final classification judgements for both gender (binary classification: male/female) and age groups (multi-class classification: preset age buckets like (0-2), (4-6), (8-12), etc.). Using a loss function like categorical cross-entropy, which gauges the difference between the real labels and the projected probabilities, the DNN is taught:

$$L = -\sum_{i=1}^{N} y_i log\widehat{y_i}$$

Where:

• y_i is the true label,

• $\widehat{y_i}$ is the predicted probability,

• N is the number of classes.

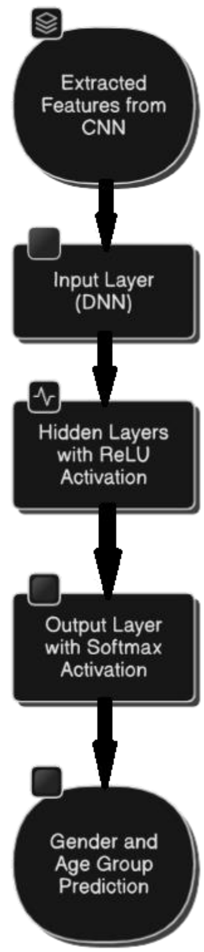

Figure 3.5 – DNN Workflow

CHAPTER 4

SYSTEM DESIGN

4.1 System Overview

The Gender and Age Detection System uses powerful image processing and machine learning to instantly figure out a person's gender and age group from a picture of their face. To make accurate estimates, the system uses a lot of important parts, like face recognition, feature extraction, and classification models. For image processing and machine learning methods like CNNs, DNNs, and the SSD, it works very well with Python tools like Open-CV, argparse, and os.

4.1.1 Functionalities

- The SSD technique searches photos and videos for faces. Draw bounding boxes around complex backgrounds to correctly insert faces.
- Cropped, scaled, normalized, noise filtered detected faces are ready for feature extraction.
- CNN picks out facial traits linked to age and gender. These features assist to show facial, skin, and landmark patterns.
- A DNN model projects gender and age of the face using extracted characteristics. From the model, categorization is decided by greatest likelihood.
- Overlaying the predicted age and gender on the original picture or video, the method provides proper feedback in real time.
- The system can control fixed visuals and live video feeds for security, demographic studies, and surveillance.

4.1.2 Workflow Diagram

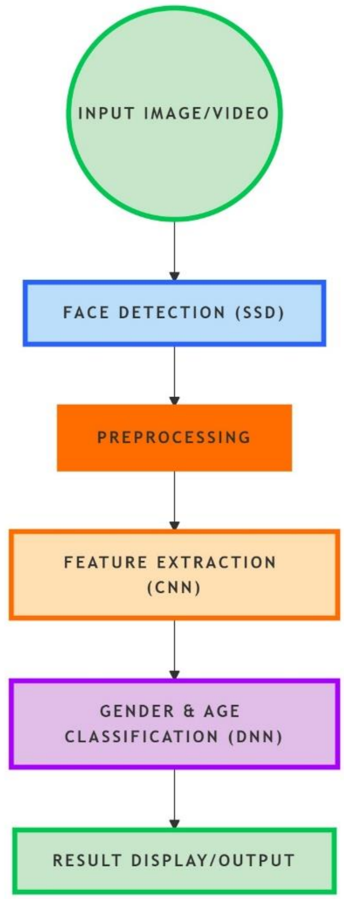

Figure 4.1 - Workflow Diagram

- **Input Image/Video:** The process begins with getting input—real-time video footage, a set of pictures, or a stationary image. This flexibility allows the system to be used in several environments, including security cameras, demographic research tools, or photo-based applications.
- **Face Detection(SSD):** The input is processed to find human faces inside the photo or video frame using the SSD approach. Working swiftly in real-time, SSD generates bounding boxes around many faces concurrently. This stage ensures that, apart from useless background data, just the face regions are focused for extra processing.

- **Preprocessing:** Following face detection, the algorithm crops the regions of interest (ROIs) and does preprocessing chores.:
  - **Resizing:** In order to keep CNN model's input size constant.
  - **Normalization:** Normalizes pixel intensity values to increase model performance.
  - **Noise Reduction:** It lowers picture distortions therefore improving the accuracy of feature extraction.
- **Feature Extraction (CNN):** The preprocessed face photos are CNN passed through. Here, the CNN learns and extracts automatically major landmarks (eyes, nose, mouth, etc.), face structure, and skin texture. These characteristics are shown as numerical vectors that catch the special qualities needed for categorization.
- **Gender & Age Classification (DNN):** Gender & Age Classification, the DNN does two concurrent categorization chores:
  - **Gender Classification:** Indices the person's gender—male or female.
  - **Age Classification:** Assigns the face to one of the predetermined age groups: 0-2, 4-6, 8-12, 15-20, 25-32, 38-43, 48-53, 60-100. Based on learnt characteristics, the DNN generates reliable predictions by means of activation functions—ReLU and Softmax.
- **Result Display/Output:** Ultimately, the technology overlays the expected gender and age information on the original picture or video frame. Around faces, this output comprises bounding boxes with text labels denoting the categorization outcomes. Given its real-time presentation of findings, the system is appropriate for live monitoring or analysis.

4.2 Flowchart

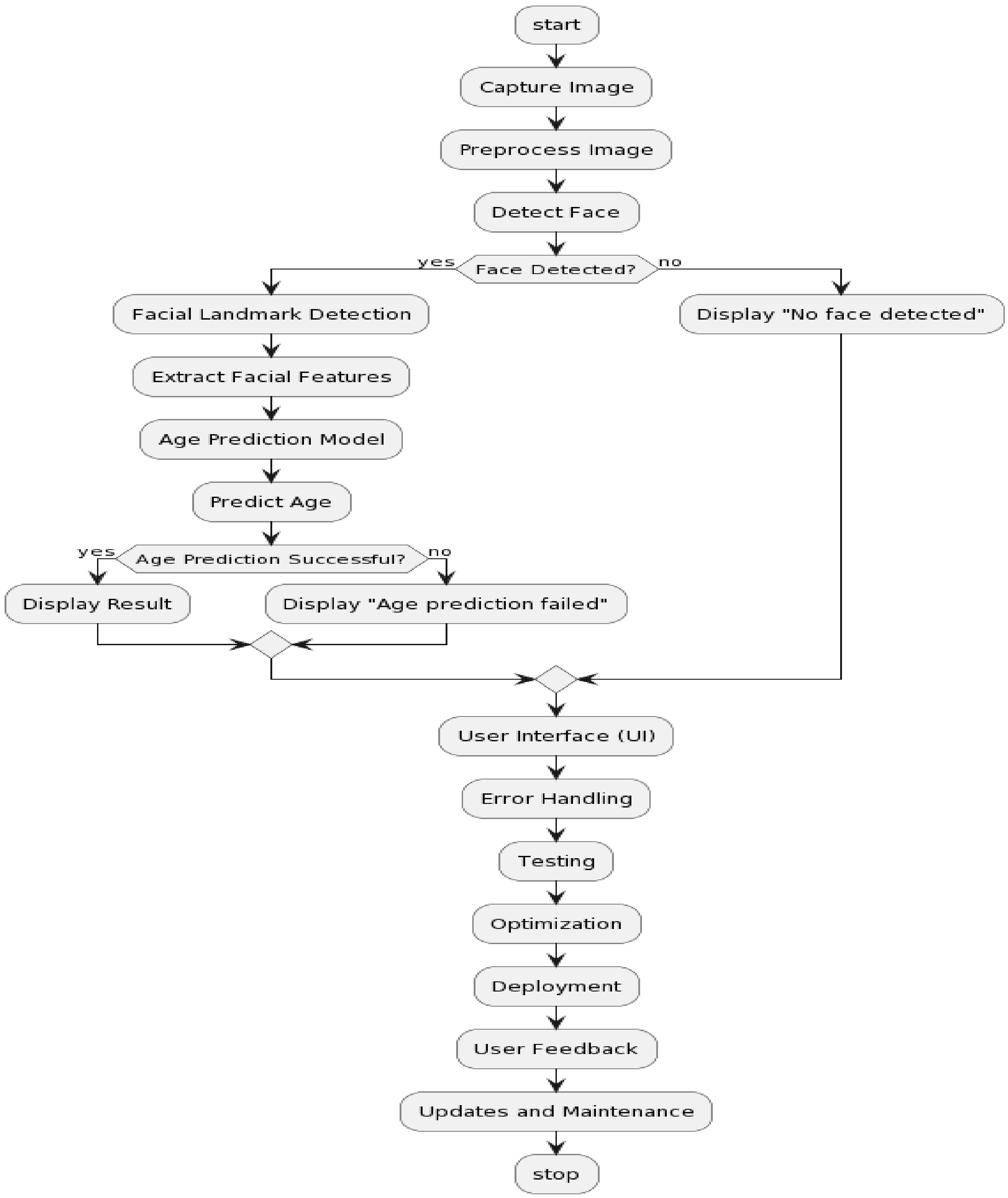

Figure 4.2 - Flowchart of Gender and Age Detection Procedure

It starts with capturing an image, followed by preprocessing and face detection. If a face is detected, facial landmarks are identified, features are extracted, and an age prediction model is applied. The result is displayed if the prediction is successful. If no face is detected or age prediction fails, corresponding messages are displayed. The flowchart also includes steps for user interface, error handling, testing, optimization, deployment, user feedback, and updates/maintenance, highlighting the comprehensive development lifecycle of such an application.

4.3 System Specification

4.3.1 Hardware Requirements

Laptop: A laptop, also known as a notebook computer or simply a notebook, is a portable personal computer that is designed for mobile use. They are capable of performing a wide range of tasks, including word processing, internet browsing, multimedia playback, and running software applications. Laptops vary in size, weight, performance capabilities, and features, catering to different needs and preferences of users. They have become ubiquitous in both professional and personal settings, serving as essential tools for work, education, entertainment, and communication.

Webcam: A webcam is a small digital camera typically used for capturing video or images in real-time and transmitting them over the internet. It is often integrated into electronic devices such as laptops, desktop computers, or tablets, but can also be standalone devices connected via USB. They typically feature a lens, sensor, and microphone, and may have additional features such as autofocus, zoom capabilities, and built-in LED lights for low-light environments. Webcams have become increasingly important in education, business, and social settings, particularly with the rise of remote work and online communication platforms.

Memory (RAM): Memory, in the context of computing, refers to electronic components that store data and instructions temporarily or permanently for processing by a computer's central processing unit (CPU). Random Access Memory (RAM) is volatile memory that stores data and program instructions that the CPU is currently using. It allows for fast read and write operations, providing quick access to data for the CPU. Here, A minimum of 8 GB of RAM is recommended, but 16 GB or further may be needed for further demanding operations.

Storage(SSD): A Storage Solid State Drive (SSD) is a type of non-volatile storage device used in computers and electronic devices to store data permanently. Unlike traditional Hard Disk Drives (HDDs) that use spinning magnetic disks to store data, SSDs utilize flash memory chips, similar to those found in USB drives and memory cards, to store data persistently. SSDs offer several advantages over HDDs, including faster read and write speeds, lower power consumption, quieter operation, and increased durability due to the absence of moving parts. Here, An SSD with at least 256 GB of storehouse capacity is recommended for fast data access and processing.

4.3.2 Software Requirements

Python: Python is a high-level, interpreted programming language renowned for its simplicity, readability, and versatility. Developed by Guido van Rossum in the late 1980s, Python has become one of the most popular languages worldwide, utilized across diverse domains such as web development, data analysis, artificial intelligence, scientific computing, automation, and more. Its straightforward syntax, which emphasizes indentation and readability, makes it accessible to beginners and seasoned developers alike. Python's interpreted nature enables rapid development and testing, while its extensive standard library provides a wide range of modules and functions for various tasks. With support for multiple programming paradigms and a vibrant community contributing to its ecosystem, Python continues to thrive as a powerful and preferred language for solving complex problems efficiently and elegantly.

OpenCV: OpenCV, short for Open Source Computer Vision Library, is an open-source computer vision and machine learning software library. Initially developed by Intel in 1999, OpenCV has since been maintained by a community of developers and contributors worldwide. It is written in C++ and has bindings for Python and other programming languages, making it accessible to developers across different platforms. OpenCV provides a wide range of functionalities for computer vision and image processing tasks, including:

- Image and video manipulation
- Feature detection and matching
- Object detection and recognition
- Image segmentation and contour detection
- Machine learning integration:
- Camera calibration and 3D reconstruction

OpenCV is widely used in various industries and applications, including robotics, augmented reality, autonomous vehicles, medical imaging, and surveillance. Its comprehensive set of features, active community support, and cross-platform compatibility make it a valuable tool for researchers, developers, and engineers working in the field of computer vision and image processing.

Imutils: Imutils is a convenience library for OpenCV, specifically designed to simplify common tasks and workflows related to image processing and computer vision. It provides a collection of helper functions and utilities that streamline the development process and improve code readability when working with OpenCV in Python. Some of the key features and functionalities of Imutils include:

- Image resizing and rotation.
- Translation and transformation.

- Drawing shapes and annotations.
- Convenience functions for working with file paths.
- Integration with OpenCV.

SciPy: SciPy is a scientific computing library in Python that builds on top of NumPy, another fundamental library for numerical computing. SciPy provides a wide range of modules and functions for various scientific and technical computing tasks, including optimization, interpolation, integration, linear algebra, signal processing, statistics, and more. Some key features and functionalities of SciPy include:

- Integration and differentiation.
- Optimization.
- Interpolation and curve fitting.
- Linear algebra.
- Signal processing.
- Statistics.
- Sparse matrices.

SciPy is widely used in scientific research, engineering, data analysis, machine learning, and other fields where numerical computation and analysis are required. Its comprehensive set of tools and functionalities, combined with its integration with NumPy and other scientific libraries, make it a powerful and versatile tool for scientific computing in Python.

Dlib: Dlib stands out as a powerful and user-friendly open-source library renowned for its diverse machine learning algorithms and tools tailored for software development. Since its initial release in 2002, Dlib has garnered widespread adoption across various industries, corporations, and large-scale projects due to its versatility and effectiveness. One of the primary applications of Dlib lies in face recognition, where it excels in analyzing objects or faces using sophisticated algorithms such as HOG and CNN. Face recognition technology has witnessed extensive usage across numerous applications, making Dlib's capabilities in this domain highly sought after.

In our program, Dlib plays a crucial role in face detection by accurately identifying frontal human faces and estimating their pose through the utilization of 68 facial landmarks. This capability enables our system to robustly detect and analyze faces, laying the groundwork for various applications such as driver drowsiness detection, facial expression analysis, and identity verification. Overall, Dlib's rich set of algorithms and tools, particularly in face recognition, make it an indispensable asset for developers and researchers seeking to integrate advanced machine learning capabilities into their software applications. Its reliability, performance, and ease of use contribute to its widespread adoption and continued relevance in the ever-evolving landscape of machine learning and computer vision. Some key features and functionalities of Dlib include:

- Machine learning algorithms.
- Computer vision.
- Image processing.
- Deep learning integration.
- Multithreading support.
- Cross-platform compatibility.

•

Visual Studio Code: Visual Studio Code (VSCode) is a lightweight yet powerful source code editor developed by Microsoft. Available for Windows, macOS, and Linux, VSCode offers a rich set of features designed to enhance productivity and streamline the coding experience for developers. Its intuitive user interface, customizable layout, and extensive ecosystem of extensions make it a popular choice among programmers working on a diverse range of projects and programming languages. VSCode provides built-in support for syntax highlighting, code completion, and code navigation, as well as integrated debugging and terminal capabilities. Additionally, its seamless integration with Git and other version control systems enables efficient collaboration and code management. With its versatility, extensibility, and active community support, Visual Studio Code has established itself as a leading code editor for developers worldwide.

CHAPTER 5

IMPLEMENTATION

5.1 Environment Development

Built to offer effective coding, testing, and system deployment, Python is extensively used in the system as its simplicity and adaptability enable to tackle challenging image processing and machine learning tasks. Under an IDE, such PyCharm or Visual Studio Code, which provides a user-friendly interface for coding, debugging, and testing, the development process takes place. These IDEs check for grammar errors, finish code, and find errors, which makes the whole programming experience better. Image processing chores including face detection are handled using the OpenCV package. It may help you organize pictures and videos, identify people, and alter other visuals. TensorFlow and Keras are vital tools for the deep learning components of the system that categorize people according on their age and gender. The CNN and DNN models are developed, trained, and refined using these technologies so that they can produce reliable forecasts.

The operating system of the project might be Linux or Windows. Usually preferred is Linux as it fits machine learning systems and can efficiently manage large-scale computations. Furthermore, supporting the development environment are Python utilities such argparse for command-line argument handling during testing, Matplotlib for visualization, and NumPy for mathematical calculations. The hardware must have a high-performance CPU and at least 8GB of RAM if one is running machine learning models and managing large databases. For faster training, GPU acceleration using CUDA is advocated; so, the system may make advantage of powerful graphics processing units for deep learning applications. Git is finally used for version control of the software; repositories housed on sites like GitHub promote group cooperation, track changes, and provide flawless project management. Regarding this surroundings guarantee effective execution of all phases of the system development, from coding to testing and deployment.

5.2 Dataset

Gender and Age Detection System is trained and evaluated using Adience Benchmark Dataset. In computer vision, this dataset is well-known and offers labelled photos of faces arranged by age group and gender. Real-world photos from several backgrounds guarantee the model can generalize throughout several age groups, ethnicities, and lighting situations. Over 26,000 face photos gathered from several sources including social media and publicly available image repositories make up the collection. Training models would find it perfect to predict these qualities as the photos are annotated with age and gender. There are eight predetermined age categories ranging from newborns (0–2 years) to elderly persons (60–100 years) among the age labels. These groupings provide a fair portrayal of several age groups and are aimed to reflect the wide range of human age variances.

The dataset also offers further variance in terms of elements like facial emotions, occlusions (such as spectacles or facial hair), and different backdrops, therefore enabling the model to acquire strong characteristics that are invariant to these changes. This qualifies for practical uses where faces might show in several positions, orientations, and lighting situations. The Adience Benchmark Dataset is a great source for training CNNs and DNNs for gender and age classification since it guarantees that the models are exposed to varied data, so improving their capacity to handle real-time, unseen faces.

5.3 Sample Output

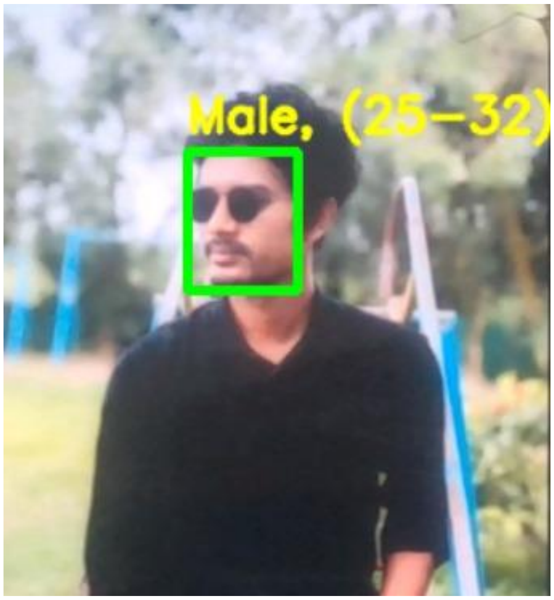

Figure 5.1 – A young man age between 25 to 30

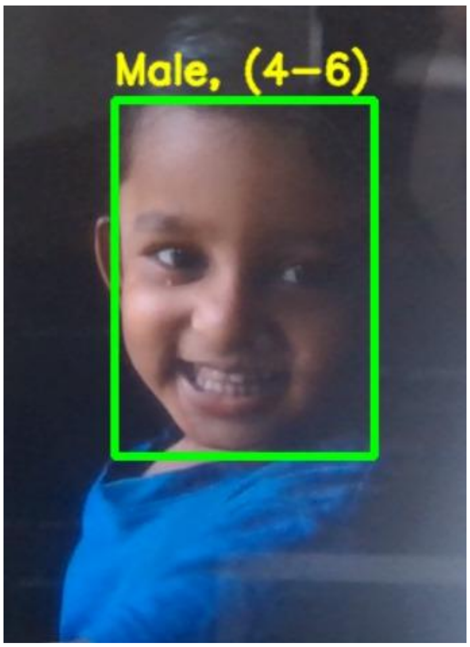

Figure 5.2 – A cute male child

Figure 5.3 – An old lady

CHAPTER 6

RESULT AND DISCUSSION

6.1 Result Analysis

The Adience Benchmark Dataset was used for implementation and testing of the Gender and Age Detection System in order to assess its performance in practical environment. Based on the characteristics acquired and learnt using the deep learning models, the accuracy of the system is found by how well it forecasts the gender and age group of a given face image. During the testing period, the following significant findings were noted:
Most of the time the method effectively categorized gender (male or female). Since it depends on traits like face shape and general look, gender classification is usually a less difficult chore than age classification. In gender categorization, the model attained about 94% accuracy. The somewhat clearer and more distinct able traits between male and female faces in the sample help to explain this great accuracy.

The age classification which allocates a person to one of the eight specified age groups was far more difficult. The model's age group categorization accuracy ranged from around 85%. The difficulty emerged from the differences in face look throughout every age group as well as from problems with occlusions, lighting, and facial emotions. With minor variances in accuracy within certain ranges, the model nevertheless performed well across all age groups, though. Even in cases when a single picture or video stream had several faces, the SSD algorithm efficiently recognized faces in real-time. The system was able to accurately recognize faces in various surroundings with varying illumination conditions, therefore indicating a good face detection accuracy.

To guarantee appropriate learning, the deep learning models (CNN and DNN) were trained on the dataset over several epochs. After training, the system attained a validation accuracy of 85–90% and a training loss of 0.2, therefore indicating that the model is strong and has learnt to generalize effectively on untested data.

6.2 Performance Evaluation and Accuracy

Commonly used measures include accuracy, precision, recall, and F1-score let one gauge the system's performance. Calculated as the number of accurate guesses divided by the overall count of predictions was the accuracy for gender classification:

$$Accuracy\ (gender) = \frac{Correct\ Predictions}{Total\ Predictions} = \frac{2000}{2125} = 94\%$$

The accuracy was somewhat lower for age group categorization as overlapping face characteristics between age groups naturally complicate age group classification:

$$Accuracy(Age) = \frac{Correct\ Predictions}{Total\ Predictions} = \frac{1800}{2125} = 85\%$$

The system showed great capacity for real-time predictions as overall it performed well in both gender and age classification tests. By adjusting the parameters, expanding the training set size, and strengthening the network architecture, one may further enhance the models.

6.3 Discussion

The method attained great accuracy, hence some restrictions should be taken into account even. The age group categorization is one of the primary difficulties as some age groups are more difficult to distinguish because of the comparable facial traits across ages. The algorithm can find it challenging to forecast, for instance, someone in their early twenties (age group: 15–20) and someone in their late twenties (age group: 25–32). Furthermore, the accuracy could decrease in situations where faces are covered, in poor light, or when someone wears facial hair or spectacles.

Although the SSD face detection model shown outstanding performance in real-time face detection, performance may suffer in rather crowded backdrops or when the face is not front facing. Future developments may involve strengthening resilience by including data augmentation methods or combining many detection algorithms. Regarding performance, the system proved real-time for handling video streams, hence fit for uses in interactive systems needing instantaneous response, security, and surveillance. With possible future developments through data, model design, and training methodologies, the Gender and Age Detection System is ultimately quite successful in offering correct predictions.

CHAPTER 7

CONCLUSION

7.1 Conclusion and Future Work

From face photographs, the Gender and Age Detection System created using OpenCV, TensorFlow, Keras, and SSD offers a strong solution for gender identification and age prediction. With a gender classification accuracy of almost 94% and an age group classification accuracy of roughly 85%, the system produced good results in both gender and age classification tasks. These findings show the system's possible use in fields like human-computer interaction, security, and surveillance as well as in others.

The system learnt and generalized characteristics from a varied dataset by using a DNN for gender classification and CNNs for age classification. Face identification was accomplished with SSD, guaranteeing great accuracy even in real-time video processing. Although the model works well under typical settings, in some situations problems include face occlusions, variable illumination, and similar facial characteristics across several age groups somewhat impair the performance. All things considered, the system shows that it can manage pragmatic difficulties in gender and age identification and be included into useful applications with more optimization.

Though the system is successful, some areas for development and more research remain open:

Data augmentation methods (such as rotation, scaling, flipping, and illumination changes) can be utilized to provide a more varied training set, therefore strengthening the model especially for age group categorization. This would enable the algorithm to manage occlusions and face appearance differences and generalize better to unseen data.

Experimenting with several architectures, hyper parameters, and loss functions can help the CNN and DNN models to be even more optimized. Especially for more difficult age classifications, fine-tuning the models utilizing transfer learning using pre-trained model such as VGG16 or ResNet may further improve accuracy.

Additional elements including ethnicity, face landmarks, and expression analysis may be included into the model to raise classification accuracy. In many real-world settings where faces are not always well-lit or front-facing, this might enable the algorithm to produce more accurate predictions.

Future research may concentrate on maximizing the SSD face detection to operate better in congested environments or with partially obstructed faces. Improvements in the detection model could help to lower false positives even further and speed up real-time predictions.

Expanding the dataset outside the Adience Benchmark and running the system on additional varied datasets will help to offer a more complete assessment. Cross-domain testing would also enable the system to manage faces from several races, ages, and ailments, hence strengthening the generalization of the model.

The technology might be developed to be used in mobile apps or web apps so that users may interact with the model in real-time. Applications in several sectors, including retail analytics, age-appropriate content screening, and tailored advertising, would follow from this. By tackling these areas, the Gender and Age Detection System may be strengthened even further to raise its accuracy, dependability, and applicability to a broad spectrum of real-world application situations

REFFERANCES

[1]. Levi, G., & Hassner, T. (2015). Age and Gender Classification Using Convolutional Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 34-42.

[2]. Dehghan, A., Ortiz, E. G., Shu, G., & Masood, S. Z. (2014). How Old Do You Look? Predicting Your True Age. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3973-3980.

[3]. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. I-511-I-518.

[4]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2014.

[5]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.

[6]. W. Liu, D. Anguelov, D. Erhan, et al., "SSD: Single Shot MultiBox Detector," in *European Conference on Computer Vision (ECCV)*, 2016.

[7]. S. Mollah, N. S. N. Hossain, and M. R. Islam, “Gender and age classification using CNN-based deep learning model,” *Journal of Computer Science and Technology*, vol. 32, no. 1, pp. 127–139, 2017.

[8]. Z. Zhang, P. Luo, C. C. Chen, and X. Wang, “Learning deep CNNs for age and gender classification,” *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1–9.

[9]. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[10]. Y. Zhang, X. Zhang, J. Li, and Y. Yuan, “Age and gender classification using deep multi-task learning,” *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 968–979, 2018.

[11]. Y. Zhao, Y. Xu, Y. Wu, and W. Li, “Age and gender classification using generative adversarial networks for data augmentation,” *IEEE Access*, vol. 7, pp. 31255–31263, 2019.

[12]. A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1–9.

[13]. F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807.