Digidetective: An Image Enhancement and Identification System

A major project report submitted in partial fulfilment of the requirement for the award of degree of

Bachelor of Technology

in

Computer Science & Engineering

Submitted by

Manav Gupta (201390)

Vipransh Agarwal (201448)

Under the guidance & supervision of

Dr. Shubham Goel

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology, Waknaghat, Solan - 173234 (India)

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: ....

Type of Document (Tick): **PhD Thesis** **M.Tech/M.Sc. Dissertation** **B.Tech./B.Sc./BBA/Other**

Name: ______________________________**Department:** _________________**Enrolment No** __________

Contact No. ______________________________**E-mail.** ____

Name of the Supervisor: ____

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): _________________________

____

____

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

- Total No. of Pages =
- Total No. of Preliminary pages =
- Total No. of pages accommodate bibliography/references =

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found Similarity Index at .... (%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

(Signature of Guide/Supervisor) **Signature of HOD**

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	**Excluded**	**Similarity Index (%)**	**Abstract & Chapters Details**	
	• All Preliminary Pages • Bibliography/Images/Quotes • 14 Words String		Word Counts	
Report Generated on			Character Counts	
		Submission ID	Page counts	
			File Size	

Candidate's Declaration

We hereby declare that the work presented in this report entitled **'Digidetective: An Image Enhancement and Identification System'** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering** submitted in the Department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology, Waknaghat is an authentic record of our own work carried out over a period from August 2023 to May 2024 under the supervision of **Dr. Shubham Goel** (Assistant Professor (Senior Grade), Department of Computer Science & Engineering and Information Technology).

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

(Student Signature with Date)
Name: Manav Gupta
Roll No.: 201390

(Student Signature with Date)
Name: Vipransh Agarwal
Roll No.: 201448

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

(Supervisor Signature with Date)
Supervisor Name: Dr. Shubham Goel
Designation: Assistant Professor (Senior Grade)

ACKNOWLEDGEMENT

Firstly, we express our heartiest thanks and gratefulness to almighty God for his divine blessings that makes it possible for us to complete the project work successfully.

We are really grateful and wish our profound indebtedness to our Supervisor **Dr. Shubham Goel (**Assistant Professor (Senior Grade), Department of CSE Jaypee University of Information Technology, Waknaghat). Deep Knowledge & keen interest of our supervisor in the field of **"Image Enhancement and Identification"** helped us to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages has made it possible for us to complete this project.

We would like to express our heartiest gratitude to **Dr. Shubham Goel** (AssistantProfessor (Senior Grade), Department of CSE), for his kind help to finish our project.

We would also generously welcome each one of those individuals who have helped us straightforwardly or in a roundabout way in making this project a win. In this unique situation, we might want to thank the various staff individuals, both educating and non-instructing, which have developed their convenient help and facilitated our undertaking.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

Manav Gupta (201390)

Vipransh Agarwal (201448)

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

LIST OF ABBREVIATIONS

1. CNN: Convolutional Neural Network
2. MTCNN: Multi-Task Cascaded Convolutional Neural Networks
3. GAN: generative adversarial network
4. SRGAN: Super-Resolution Generative Adversarial Network
5. PCA: Principal component analysis
6. GFPGAN: Generative Facial Prior Generative Adversarial Networks

Abstract

In an era where security and personalized access control play an important role in diverse domains, the integration of face detection and identification systems has become imperative. This project focuses on the development of an advanced facial recognition system, combining state-of-the-art techniques in face detection and identification to address the growing demand for secure and seamless authentication processes.

The project starts with the implementation of an accurate face detection module, consisting of deep learning architectures such as convolutional neural networks (CNNs). The selected model is trained on diverse datasets to ensure adaptability to various environmental conditions, lighting scenarios, and facial orientations. Real-time face detection capabilities are given priority to enhance the system's usability in dynamic and time-sensitive applications.

Building upon the face detection module, the project integrates facial identification to create a reliable biometric authentication system. This includes the creation of a diverse facial database and the utilization of advanced feature extraction techniques. Famous algorithms, including deep face recognition models, are employed to accurately match detected faces with stored identities, which ensures a high level of precision and security.

Ethical considerations are an important, aspect of the project, and privacy concerns are addressed through the use of robust data protection measures. The system is designed with privacy-preserving features, such as anonymization techniques and secure storage protocols, to mitigate the risks associated with unauthorized access and misuse of facial data.

The effectiveness of the proposed face detection and identification system is rigorously tested and evaluated through extensive testing in controlled environments and real-world scenarios. Performance metrics, including accuracy, speed, and false positive rates, are measured and analysed to validate the system's reliability and efficiency.

Chapter 1 Introduction

1.1 Introduction

The increasing advancements that have happened in our environment have seen this turn out as one of the strongest allies during war against terrorism. Technology-assisted law enforcement innovations have been and are currently being developed for reducing crime, enhancing security, and improving living. It is also worth noting that technology has played a great role in improving quality data analytics services, and enhanced surveillance systems for crime detection as well as arresting of those involved. Changes have also affected how crime happens, especially in how police use CCTV footage to catch criminals. We aim to keep such mental defectives as low as possible among us.

Our principle in this project involves getting images out of the surveillance videos and looking for an image that shows the face features of the possible suspect clearly. The enhanced frame will then be combined and averaged with the surrounding frames, yielding a high-quality image from which the faces will be cropped, enhanced and compared to stored records. This project examines different possible ways of improving picture quality, such as histogram equalization, contrast stretching, spatial filtering, and super-resolution. These methods are very important for overcoming problems of noise, poor resolution, and unfavourable illumination. State-of-the-art algorithms including deep learning models, Haar cascade, facial landmark detection, and facial embeddings provide both face detection and face recognition simultaneously. These improvements allow for pinpointing of personalities on image, including when it is under unfavourable conditions of movement.

Image enhancement combined with facial detection can be applied effectively in security and surveillance, healthcare, entertainment, as well as others. Insite of that, such concerns are raised on matters related to privacy, adversarial attacks, and real-time processing. Therefore, the project focuses on the ongoing issues and future trends and the importance of solid and moral answers. This multidisciplinary attempt aims toward a greater comprehension of visual information for a more informed computer vision, concluding in enhanced facial recognition in the increasingly universal connectivity.

1.2 Problem Statement

In the context of improving public safety and regulation enforcement, the primary goal of this project is to expand an advanced facial recognition machine designed mainly for the identity of people with criminal data inside video CCTV footage. The challenges to be addressed are combined in different possible ways related to the precise requirements of forensic facial evaluation and criminal identity:

1.2.1 Forensic-Quality Facial Detection:

Create a facial detection module optimized for forensic packages, able to appropriately figure out faces inside video CCTV footage, even in tough conditions together with low decision, varying angles, and occasional lights. The gadget has to prioritize sensitivity in detecting faces of interest.

1.2.2 Facial Feature Enhancement For Identification:

Superior facial enhancements and applications particularly designed for forensics purposes. Improve the readability and knowledge level of facial photographs to easily identify individuals with criminal records.

1.2.3 Integration With Criminal Databases:

Connect and integrate face recognition software across to the criminal database and match any existing photos in real-time when necessary. Therefore, it is important that a design for system that can effectively manage very large databases has to be adopted.

The proposed assignment intends to notably contribute to the field of forensic facial recognition by giving law enforcement agencies an effective and powerful tool for quick and accurate identification of wanted individuals whose images are contained in the video-CCTV footage. It ensures privacy as well as legal enforcements and ensures ethics requirements and norms of the society.

1.3 Objectives

The objectives of a project are focused on improving facial identification of criminals from video CCTV footage that can be framed to address the specific challenges and goals of the project. Here are potential objectives for such a project:

- Create a Robust Facial Detection Algorithm:
  Create a face detection algorithm that can identify faces in CCTV footage in variety of conditions and in low-resolution, multi-angle, and dimly lit environments.

- Put Forensic-Quality Facial Enhancement Techniques into Practice:
  Use protocols created and designed especially for forensic programs to improve the readability and content of images, which will increase the algorithm's capacity to identify people with criminal histories.

- Integrate with Criminal Databases for Real-Time Matching:
  Establish seamless integration with criminal databases to allow actual time matching of detected faces against an extensive database of regarded offenders, facilitating enhanced identification all through surveillance.

- Ensure Privacy and Legal Compliance:
  Implement privacy-keeping capabilities, such as face anonymization at some point of non-identity eventualities and adhere strictly to legal frameworks and policies governing the use of facial popularity era in criminal investigations.

- Integrate with Existing Surveillance Infrastructure:
  Design the system for seamless integration with numerous video surveillance infrastructure, making sure compatibility with diverse CCTV systems and protocols for huge applicability.

- Enhance System Reliability and Redundancy:
  Implement redundant systems and failover mechanisms to make certain non-stop operation, enabling the reliability of the facial recognition device, especially in essential regulation enforcement situations.

By setting those targets, the venture goals to create a facial identity system that is not only most effective, technically robust however additionally aligned with legal, moral, and operational considerations which are important inside the context of criminal identification from video CCTV pictures.

1.4 Significance And Motivation of The Project Work

Enhancing Surveillance Systems: Improving picture enhancement and face detection technologies is essential for boosting the talents of surveillance systems. High-quality, better snap shots and correct face detection can aid in identifying individuals and capacity threats in safety photos.

- When preventing unauthorized access: Facial recognition with high image accuracy can help prevent unauthorized access to secure areas or systems. It is important to identify individuals attempting to breach security measures.

- Crime prevention: High quality imaging and accurate facial recognition play an important role in crime prevention and investigation. Their motivation is to develop tools to help law enforcement identify criminal suspects and solve crimes.

- Counterterrorism and National Security: Counterterrorism and National Security: Image accuracy and facial recognition technologies are critical to national security counterterrorism efforts. It can be linked to protecting national security by improving threat detection capabilities.

1.5 Organization Of Project Report

We are developing An Image Quality Enhancement and Person Identification System for which we first went through different research paper to gain more information regarding our project. From the information gained we started making a dataset of different CCTV Footage. We specifically took those videos where the person is not looking into the camera and video quality is not good. This was done to make out mode more realistic.

We started by extracting frames from the CCTV footage in order to find differences and similarities between different scenes. Based on the information found we used MTCNN model for facial detection.

After detecting the face, we aimed to improve the clarity of face. For this we needed a model that can detect and reconstruct that person's face. For this we used PCA, Eigenvalues and Eigenvectors of that person's face.

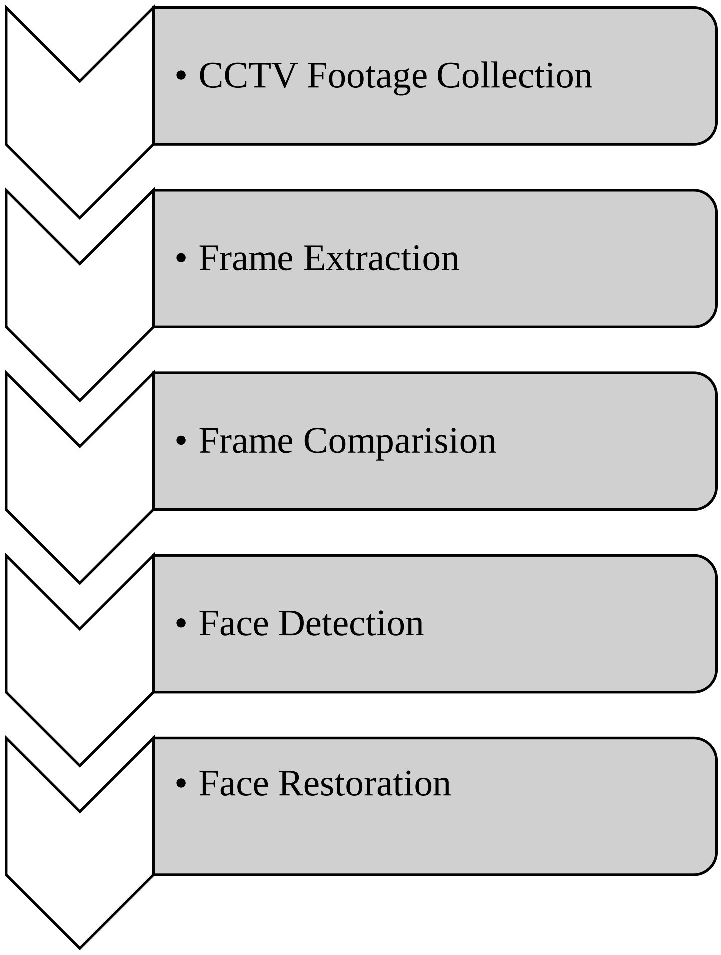


Figure 1.1: Flow Chart

Chapter-2 Literature Survey

Overview

The first article we read was Extreme Low-Light Image Enhancement for Surveillance Cameras Using Attention U-Net[2]. This article offers with extremely-low mild photo enhancement for surveillance cameras using the Attention U-Net. It discusses the demanding situations of low-mild images and the want for brand new techniques. Its opinions present strategies and their boundaries.

The article then proposes a brand-new Attention U-Net-based approach that outperforms present techniques. The Attention U-Net approach is a deep mastering model that makes use of the eye technique to consciousness at the most important elements of a photograph. The article concludes with a dialogue of the applicability of the proposed technique. The proposed Attention U-Net approach has the ability for use in a lot of programs, together with surveillance cameras, night vision cameras, and scientific imaging.

Another paper titled CCTV Surveillance Camera's Image Resolution Enhancement using SRGAN[11] introduced new method of image enhancement. This paper proposes a new technique for CCTV picture enhancement the usage of SRGAN. SRGAN is a deep learning model that can be used to improve the decision and reduce noise in pix. The authors educated their SRGAN version on a dataset of CCTV photos. The results display that the model can drastically improve the quality of CCTV pics. Three The authors evaluated their version on quite a few CCTV snap shots, inclusive of pics of human beings, automobiles, and scenes. The effects show that the version can enhance the resolution of CCTV pix by way of up to 4x and decrease noise by means of as much as 50.

Although it had some limitations such as the proposed method is not optimized for Video SR in real time and the processed images have high peak SNR but lack high frequency texture and details.

The article Comparative Analysis of CCTV Video Image Processing Techniques and Application[11] is set developing video photographs for CCTV programs. It discusses the challenges of implementing CCTV video and approaches to conquer it. Highlights of this article include the numerous components of CCTV video programs and the particular procedures used at every stage. It also mentioned some limitations such as accuracy of the processing can be affected by the placement of the cameras and the lighting conditions.

Another paper reviewed the image enhancement techniques[1]. This is an editorial on photo enhancement strategies. It discusses what photo enhancement is and why it's miles critical. It also goes into detail approximately exceptional methods of enhancing the image for exclusive packages. Some crucial factors from this text are that image enhancement may be used to enhance the visible appearance of a photograph and there are numerous extraordinary approaches to apply it in unique programs.

It also discusses some limitations of these algorithms such as inability to accurately detect objects in bad weather or when they are obscured by other objects and accuracy of the processing can be affected by the placement of the camera and the angle of the line of sight.

A Novel Key-Frame Extraction Method for Both Video Summary and Video Indexing" through Yu et al. (2014)[9] suggest a new key-body extraction method that can be used for video summarization and video indexing.

The proposed method first segments the video into subshots based totally on semantic shape the use of a dynamic tele segmentation functionality algorithm. Then, it makes use of single price decomposition (SVD) to extract keyframes from every subshot. But this method also had some limitations such as Computational complexity as the recommended approach requires significant processing resources, particularly for lengthy films. This is the result of using the dynamic distance separability set of criteria in conjunction with the SVD technique. Another limitation was sensitivity to hyperparameters as the proposed technique requires a large number of hyperparameters, a wide range of clusters, and an interframe distinction threshold in order to function. The total performance of the proposed approach is highly dependent on the selection of those hyperparameters. It also didn't showed robustness to occlusion. The recommended method might not be robust to occlusion if the occluded items contain semantic meaning. - Performance on low-gross box office receipts: These measures how well the recommended approach works across the board on low-gross movies, or movies with audio or compression issues.

Effective video face recognition primarily based on body choice and best evaluation[8] This article with the aid of Kharchevnikova and Savchenko (2021) offers a novel method to video face reconstruction based on body selection and great evaluation recommendations for. The proposed technique can appropriately come across faces in films even supposing faces are blocked or the video great is negative.

This paper also mentioned some of its limitations such as computational complexity as this proposed method is computationally steeply-priced, especially for lengthy motion pictures. This is because of using deep mastering-based totally fashions for coverage selection and optimization assessment. Another one was sensitivity to hyperparameters as the proposed method relies on some of hyperparameters together with the frame choice threshold and the quantity of frames to be selected. The performance of the proposed method is sensitive to the selection of those hyperparameters. Also, performance on low-quality videos the effectiveness of the proposed approach is not obvious for low-first-class movies with noise or compressed features. The proposed method may not be strong to occlusion, particularly if the occluded gadgets are semantically vital.

For the frame selection process, a paper titled Quality Based Frame Selection for Face Clustering in News Video[3] mentioned some information. The authors measured the temporal distribution of the faces to ensure that the selected information was sometimes distributed throughout the sequence. The authors evaluated their method on news video database and found that not selecting feature-based frames significantly improves face set performance compared to the original method but, quality font selection for face sets in news video Bad on the paper presents a novel and the method is based on a combination of low- and high-level features, and considers face distribution over time. The authors evaluated their method on a news video database and found that it significantly improved face-clustering performance. This paper also mentioned some of its limitations such as computational complexity as this proposed method is computationally steeply-priced, especially for lengthy motion pictures. This is because of using deep mastering-based totally fashions for coverage selection and optimization assessment. Another one was sensitivity to hyperparameters as the proposed method relies on some of hyperparameters together with the frame choice threshold and the quantity of frames to be selected. The performance of the proposed method is sensitive to the selection of those hyperparameters. Also, performance on low-quality videos the effectiveness of the proposed approach is not obvious for low-first-class movies with noise or compressed features. The proposed method may not be strong to occlusion, particularly if the occluded gadgets are semantically vital.

An improved face recognition algorithm and its application in attendance management system[4] a novel sturdy face identity technique for modifications in lights and facial developments is proposed in this observe. Deep studying and local binary models (LBPs) form the inspiration of the approach. Local functions are extracted from the face photo using LBP functions, while worldwide features are extracted from the face picture the usage of deep getting to know capabilities. Following characteristic extraction, the Support Vector Machine (SVM) classifier is trained the usage of a mixture of the functions. But instead of all of the functions it has some limitations such as the proposed algorithm is computationally expensive, not robust to occlusion and the performance of this method is not clear for low quality noisy videos.

A framework for face recognition and tracking in human-robotic interaction (HRI) is offered in Face Recognition and Tracking Framework for Human–Robot Interaction[7] look at. A deep studying model trained for in-the-moment insights and management serves as the inspiration for the framework. Extreme occasions such as shifts in mild, temper, and obstruction also can be treated by means of the device.

Using HRI video datasets for trying out, the machine validated notable accuracy and robustness. But regardless it has some limitations such as the proposed algorithm is computationally expensive, not robust to occlusion and the performance of this method is not clear for low quality noisy videos. Also, the framework has not been evaluated on a large dataset of HRI videos with different types of content.

CenterFace[12] is a one-tier face recognition and alignment system based on a deep learning model. The model is trained to predict the continuity on the face, and five markers from the focal point (two eyes, nose, two corners of the face) The model is also trained to predict the confidence score of internal face recognition.

It also has some limitations like computational complexity as this proposed method is computationally steeply-priced, especially for lengthy motion pictures. Another one was sensitivity to hyperparameters. CenterFace is sensitive to the choice of hyperparameters, such as the learning rate and the number of epochs to train the model. The performance of the proposed method is sensitive to the selection of those hyperparameters. Also, the performance of CenterFace on low-quality images, such as images with noise or compression artifacts, is not clear.

Real Time Face Recognition System with Deep Residual Network and Knn[5] introduces One of the most popular methods of entry, security and surveillance is facial recognition. Extracting artifacts from face images are the most common methods for traditional face recognition. however, these processes are generally sensitive to environmental changes and light exposure.

Deep learning has become a powerful technique in recent years for image recognition. Deep learning models are more resilient to changes in data entry because they can extract complex features from the data. The authors present a real-time face detection algorithm that combines K-Nearest Neighbours (KNN) classification for face detection and deep residual network (DRN) for feature extraction. KNN is a simple and efficient classifier that may be used to categorise new records factors primarily based on similarity with previously identified facts factors. The authors evaluated their proposed method on benchmark datasets: the CMU Multi-PIE dataset and the FERB dataset.

On the CMU Multi-PIE facts set, their proposed technique performed an accuracy of 90.2%. This is extensively better than the accuracy of traditional techniques, which generally obtain an accuracy of 70-80%. On the FERB dataset, their proposed method performed an accuracy of 88.5%. This is also significantly better than the accuracy of conventional methods, which usually obtain an accuracy of 60-70%. The limitations of the paper are that it only considers live streaming on surveillance systems and that it only uses a KNN classification model.

Face recognition using multitasking cascading convolutional networks[10]. This article discusses the usage of convolutional neural networks (CNNs) for face recognition. CNNs are a type of deep studying paradigm this is in particular nicely ideal for visualisation obligations.

CNNs excel in face recognition due to their potential to recognize hierarchical systems from visible information. Features are extracted from many pix, and the representations of facial capabilities are progressively delicate. This hierarchical method allows CNNs to capture complicated variations in human faces, such as facial expressions, lights, and expression.

A traditional CNN scheme for face reputation includes convolutional layers, pooling layers, and absolutely linked layers. Convolutional layers cast off capabilities from the enter image, at the same time as pool layers reduce the dimensionality of function maps. Fully overlapping layers integrate filtered features to categorise every part of the picture as containing a face or no longer.

CNNs are skilled on huge statistics sets of categorised snap shots, in which each photo is annotated with the locations of sure facial features. During schooling, CNN learns to regulate its weights to lessen the mistake between its predictions and genuine scores. This phenomenon, called historical past propagation, allows the CNN to perceive complicated shapes that distinguish faces from other features in photos.

The authors suggest a multitasking cascading CNN architecture for face reputation, which they declare outperforms other cutting-edge techniques in various conditions The authors' proposed framework consists of CNNs: face popularity and face a they see the oath. The face popularity community controls the faces inside the image, whilst the face recognition network controls the two cascade networks of acknowledged faces.

But it also has some limitations such as overfitting. CNN versions that learn school data well but neglect to incorporate further information are prone to overfitting. Because faces can be better recognized under particular lighting conditions, positions, and gestures, this could be troublesome for facial recognition algorithms. Another one is robustness. Adversary assaults that are skilfully constructed to deceive the system into making mistakes can compromise facial recognition systems. For example, an attacker might attempt to fool the facial recognition technology into believing they are someone else, that they are wearing a mask, or anything else. Also, it has privacy concerns A lot of privacy issues are made worse by facial recognition software. Such devices run the risk of being exploited, for instance, to spy on and monitor adversaries.

Research On Current Situation Of 3d Face Reconstruction Based On 3d Morphable Models by Weirong Hu, Yu Li, And Xuesong Wang, Published In The Journal Of Physics: Conference Series In 2021 [14].

This paper discusses 3DMM. 3D morphable model, also called as 3DMM, is the mathematical formulation of the 3D shape and textures of human face. Powerful face analyser/maker, used in some applications like facial recognition, facial animation, virtual reality, etc.

3DMMs are usually generated using the fit of a mathematical model to a collection of thousands of three-dimensional images representing different faces. A new 3D face like the face of the training data is then created using the model. Shapes of faces can be created in the three-dimensional mode mimeograph (3DMMs) and those existing can be measured.

3DMMs are more superior than other forecasting methods like photos, images, and points. In contrast with pictures two dimensional or points, 3DMMs are slimmer, yet they acquire more data. In addition, 3DMMs have higher levels of transparency and provide opportunity for developing larger variety of angles.

But there are some limitations also like one limitation is that 3DMMs are typically built from a limited number of training samples, which can restrict their ability to generalize to new faces. Another limitation is that 3DMMs are often sensitive to noise and artifacts in the input images.

ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks by Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, Xiaoou Tang.[15]

This article talks about the SISR technique that is aimed to amplify single picture quality. It outlines GAN as the tool to help them better their visual feature of SISR. The authors suggest a Super-Resolution Generative Adversarial Network model with Enhanced up (ERGAN) Parameters. ESRGAN possesses the following recent innovations that are considered better than the current SRGANs.

Among other techniques are the new network's architecture, perceptual loss with enhancements, and a relativistic discriminator. The ESRGAN solution beyond SRGAN on the PIRM-SR competition also gives improved visual quality. It might be computationally expensive compared to other simpler models, and training ESRGAN from scratch requires significant computational resources.

Image Super-Resolution Using Very Deep Residual Channel Attention Networks by Yulun Zhang.[22]

This paper develops RCAN (Residual Channel Attention Networks) a novel deep network structure to high-quality image super-resolution. RCAN imploys a technique that combines residual learning blocks and channel attention strategy to have increased accuracy around the edges of the image. This method is shown to generally outperform most mature methods on the datasets used, including the self-ensemble technique. Most importantly, the use of their sentinels comes to the fore front even when the images require larger scaling factors. But it is difficult to train this model as compared to other models. Also, CNN-based method treats channel-wise features equally that can hinder better discriminative ability for different features.

Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks by Wei-Sheng Lai.

The paper `A Novel Technique of Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks`[23] reports a breakthrough method for improving the resolution of images. In this work the researchers propose the Deep Laplacian Pyramid Networks (LapSRN), a new network structure with the focus on precise and fast generation of high-resolution images from the low-resolution ones. This method merges the strength of Laplacian pyramid structure and convolutional neural networks (CNNs) at the same time for speeding up the computation and incredible outcomes. With the help of graph representations of images in a hierarchical fashion LapSRN was able to achieve both global and local image features extraction and therefore the super resolution generation could be done efficiently. Extensive experiments shows that LapSRN has faster speed and higher accuracy performance than existing super-resolution methods, which might be a reasonable candidate for image resolution enhancement. But it comes with some limitations also like it may not achieve the same level of performance as more complex models like ESRGAN or RCAN. Its performance depends on dataset and various real-world scenarios.

The evaluation of the method that was proposed is based on the most common factors like PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index). Perhaps these indices are employed, yet they may not be the most accurate ones in terms of perceptual quality. Besides considerations of more complex perceptual qualities, or user studies could bring the method evaluation to a new level.

Reconstructing High Resolution ESM Data Through a Novel Fast Super Resolution Convolutional Neural Network (FSRCNN) By Linsey S. Passarella. [24]. This research explores a brand-new use of an accelerated super-resolution convolutional neural network (FSRCNN) for the reducing of earth system model (ESM) simulation. Unlike previous super-resolution methods, FSRCNN keeps the same input feature dimension sizes as the low-resolution input; therefore, it requires less convolutional layers, resulting in less over-smoothing and savings of the computational cost. The authors reshape FSRCNN because of the improvement of reconstruction of ESM data, which is called FSRCNN-ESM. They use the month-to-month average model outputs of surface parameters over North America obtained from the USA Department of Energy Exascale Earth System Model, running on the climatological control simulation.

As demonstrated by the research by utilising of the pairs of high resolution and coarsened low resolution images for training and evaluation they show that FSRCNN-ESM out performs FSRCNN and other super-resolution methods in the process of reconstructing high-resolution images. Besides, the effectiveness of modelling it at finer spatial scale features is achieved with higher accuracy of surface temperature, surface radiative fluxes, and precipitation.

A major shortcoming of the FSRCNN-ESM is that, the robustness of the FSRCNN-ESM to variations in input data characteristics, such as different spatial resolutions or temporal scales, is not thoroughly evaluated.

Analysing the model's performance within a variety of circumstances ensures its suitability for different settings. The studies concentrate on North America and the shows FSRCNN-ESM is effective under the corresponding climatic conditions and topography. Furthermore, the performance of the algorithm in different geographical locations might change for the weather patterns or topography vary. Assessing its effects in other regions may serve as a contributing factor in reassuring its generalizability.

Chapter-3 System Development

3.1 Requirements And Analysis

3.1.1 System Requirements:

3.1.1.1 Hardware Requirements:

- High-performance servers.
- Compatibility with existing CCTV camera hardware.

3.1.1.2 Software Requirements:

- Face detection and identification algorithms.
- Database management system for criminal records.
- Integration with video surveillance software.

3.1.1.3 Network Requirements:

- Secure communication.
- Efficient Bandwidth usage.

3.1.2 Functional Requirements:

3.1.2.1 Facial Detection:

- Accurate detection of faces in CCTV footage.
- Capability various conditions for example lightning conditions.

3.1.2.2 Facial Enhancement:

- Implementation of image enhancement.
- Clarity improvement for reliable identification.

3.1.2.3 Database Matching:

- Seamless integration with criminal databases.
- Comparison with images in existing data base and criminal records.
- Effective management of giant database systems.

3.1.2.4 Privacy Preservation:

- Face anonymization in “non-identification” scenes.
- Following legal requirements in ensuring privacy.

These comprehensive needs assessment also lays the foundation for designing an optimal facial recognition system for criminal detection from video CCTV This project incorporates technical, operational, legal and ethical considerations to ensure effective implementation and they do it responsibly.

3.2 Project Design and Architecture

3.2.1 Modular Architecture:

The system is structured with modular components to enhance flexibility and scalability. The key modules include:

3.2.1.1 Facial Detection Module:

- Utilizes advanced algorithms for accurate detection of faces in video CCTV footage.
- Handles variations in lighting, angles, and resolution.
- Provides real-time face localization.

3.2.1.2 Facial Enhancement Module:

- Implements forensic-quality image enhancement techniques.
- Enhances clarity and details of facial features for improved identification.
- Works seamlessly with the facial detection module

3.2.1.3 Database Matching Module:

- Establishes a secure and seamless connection with criminal databases.
- Enables matching of detected faces against an extensive repository of known offenders.
- Supports efficient handling of large-scale databases.

3.3 Data Preparation

In our foresight project, we are committed to combining college CCTV data with student profiles for rigorous testing and accreditation. Data generation requires careful tracking of moral issues, making sure compliance with confidentiality policies and acquiring knowledgeable consent from contributors College CCTV photos is a valuable aid, which allows us simulate real studies surroundings in an academic setting. This dataset is cautiously generated from a whole lot of eventualities, capturing a wide variety of lighting fixtures situations, angles, and crowd levels, and carefully mixed to create a comprehensive check dataset that faithfully reflects the complexity of the college surroundings. This aggregate of information permits us to test the performance of the system in terms of efficaciously identifying college students and ability protection risks in a college putting by incorporating real university systems into our checking out information, we seek to enhance face seeing the performance and accuracy of the gadget improved Yet now not best enhances safety but keeps the highest moral requirements.

3.4 Implementation

Our project follows the following steps to obtain the desired output:

1. Obtaining frames with sufficient amount of facial information
2. Obtaining facial information from these frames
3. Enhancing these images for better Identification
4. Matching the enhanced facial images with the preexisting facial database

3.4.1 Frame Extraction

For obtaining the frames we will follow the following steps:

- Sampling Rate: Decide on a sampling rate or interval (e.g., every 1 second or every few frames) for frame extraction. This prevents an excessive number of frames and focuses on key moments.

```python
import cv2

def FrameCapture(path):
    vidObj = cv2.VideoCapture(path)
    count = 0
    success = 1

    while success:
        success, image = vidObj.read()
        cv2.imwrite("frame%d.jpg" % count, image)
        count += 1

FrameCapture("C:\\Users\\vipvi\\Desktop\\Major\\Code\\data\\video1.mp4")
```

Figure 3.1: Frame Extraction Code

• Key Frame Selection:

- Facial Clarity: Rank frames based on the clarity and visibility of facial features. This can be determined using image quality metrics or by evaluating the face detection confidence score.

- Pose and Expression: Prioritize frames that capture different poses and expressions. This diversity can be valuable for various applications.

- Lighting Conditions: Frames with adequate lighting conditions that minimize shadows and overexposure are preferred.

• We Will compare the frames based on the following factors:

- If two pixels have the same intensity in both frames: the absolute difference between those pixels will be zero, and the resulting interframe will be black at that point.

- If two pixels have different intensities in two frames: the absolute difference between those pixels will have a positive value, and the resulting difference frame will have a brighter position at that point, indicating change or movement at between frames.

```python
import cv2
import os

frame_directory = 'frames/'

frame_files = sorted([f for f in os.listdir(frame_directory) if f.endswith('.jpg')])

if len(frame_files) < 2:
    print("Error: There are not enough frames for comparison.")
    exit()

prev_frame = cv2.imread(os.path.join(frame_directory, frame_files[0]))

for frame_file in frame_files[1:]:
    curr_frame = cv2.imread(os.path.join(frame_directory, frame_file))
    frame_diff = cv2.absdiff(prev_frame, curr_frame)
    cv2.imshow('Frame Difference', frame_diff)
    prev_frame = curr_frame
    if cv2.waitKey(30) & 0xFF == ord('q'):
        break

cv2.destroyAllWindows()
```

Figure 3.2: Frame Comparison Code

3.4.2 Facial Information Extraction

For obtaining the facial information we will follow the following steps:

- Pre-processing:
  - Resize frames to a consistent resolution to ensure uniformity.

- Face Detection:

  Utilize a facial detection algorithm (MTCNN) to locate faces within each frame. MTCNN, or Multi-task Cascaded Convolutional Networks is a state-of-the-art deep learning algorithm designed for face recognition, facial landmark localization, which consists of three cascade networks, MTCNN uses a horizontal one, and differs in face recognition efficiency at in terms of dimensions (Proposal Network)] while providing proposals, the second one refines these areas [R-Net (Refinement Network)], and the third one performs face landmark localization [O-Net (Output Network)].

  P-Net, R-Net, and O-Net are three components of Multi-Task Cascaded Convolutional Networks (MTCNN), a deep learning framework designed for face recognition and facial landmark localization

Each stage in MTCNN serves a specific purpose in the hierarchical detection process:

1. **P-Net (Proposal Network):**
   - In the first step, the desired facial areas are proposed in a drawing. P-Net uses a fully variable mesh to optimally find an input image of possible faces, resulting in bounding box shapes and associated confidence scores This step prioritizes and checks neighbourhoods for possible faces in the middle

2. **R-Net (Refinement Network):**
   - In the second stage, the R-Net resolves the facial parameters generated by the P-Net. Accordingly, the first bounding detection box will determine the regions and then set the coordinates which will be the candidate points. R-Net may help the credibility of the people by reducing the upsetting numbers this time with less false positives.

3. **O-Net (Output Network):**
   - The final section, O-Net, specializes in localizing facial landmarks and further remodelling seemed faces. O-Net assessments key facial markers together with eyes, nose and mouth in boundary packing containers. This phase improves the accuracy of face reputation and gives additional facts for subsequent face evaluation duties.

However, after the MTCNN method above, which is widely used, even the real-world practice contains the problem caused by money, scale, and lighting, while the model has been distributed everywhere. This is a tool which implements advantage in the realm of computer vision, especially for facial recognition and analytics aiding to decide accurate and fast discriminate features in difficult images.

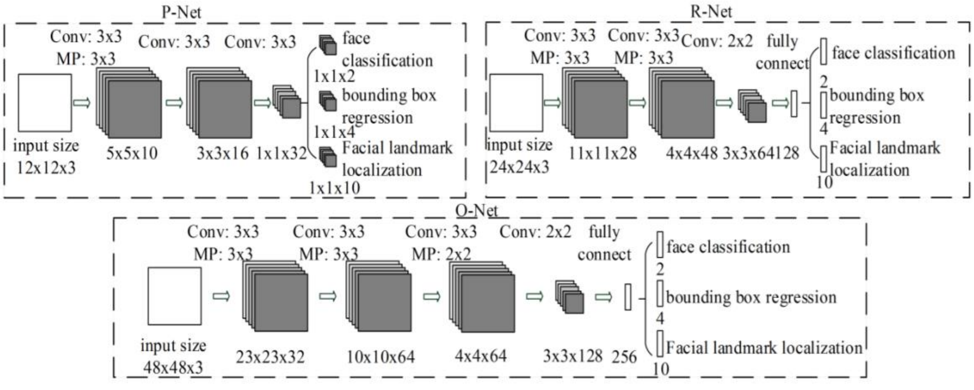


Figure 3.3: MTCNN

```python
import cv2
from mtcnn import MTCNN
import os

detector = MTCNN()

frame_directory = './data/frames/eve/'

frame_files = sorted([f for f in os.listdir(frame_directory) if f.endswith('.jpg')])

def save_face_roi(event, x, y, flags, param):
    global frame, faces
    if event == cv2.EVENT_LBUTTONDOWN:
        for face in faces:
            x1, y1, width, height = face['box']
            x2, y2 = x1 + width, y1 + height
            x1 -= 20
            y1 -= 20
            x2 += 20
            y2 += 20
            if x1 <= x <= x2 and y1 <= y <= y2:
                face_roi = frame[max(0, y1):min(frame.shape[0], y2), max(0, x1):min(frame.shape[1], x2)]
                resized_face_roi = cv2.resize(face_roi, (200, 200), interpolation=cv2.INTER_CUBIC)
                cv2.imwrite('./face_roi.jpg', resized_face_roi)
                print("Face ROI saved as 'face_roi.jpg'")
                break

for frame_file in frame_files:
    frame = cv2.imread(os.path.join(frame_directory, frame_file))
    faces = detector.detect_faces(frame)
    for face in faces:
        x, y, width, height = face['box']
        x2, y2 = x + width, y + height
        x -= 20
        y -= 20
        x2 += 20
        y2 += 20
        cv2.rectangle(frame, (max(0, x), max(0, y)), (min(frame.shape[1], x2), min(frame.shape[0], y2)), (0, 255, 0), 2)
    cv2.imshow('Face Detection', frame)
    cv2.setMouseCallback('Face Detection', save_face_roi)
    if cv2.waitKey(30) & 0xFF == ord('q'):
        break

cv2.destroyAllWindows()
```

Figure 3.4: MTCNN Code

The code above is a Python program that uses a library called MTCNN to detect faces in images and save the detected faces to a file. Below is a breakdown of the code:

Imports:

- cv2: OpenCV is a library used for computer vision tasks like image processing and video analysis.
- from mtcnn import MTCNN: Imports the MTCNN class from the mtcnn library, which is a pre-trained convolutional neural network model for face detection.
- import so: This library is used for interacting with the operating system, such as file paths.

Global Variables:

- detector: An instance of the MTCNN class is created to be used for face detection throughout the program.

Function Definitions:

- save_face_roi: This function is called when the user clicks the left mouse button on the image window. It iterates through the detected faces and checks if the click coordinates fall within the bounding box of a face. If so, it extracts the region of interest (ROI) around the face, resizes it, and saves it as a new image file.

Main Script:

1. **Load Frames:**

   - frame_directory: This variable stores the directory path where the images (frames) are stored.
   - frame_files: This list stores the filenames of the images in the frame directory.

2. **Loop through Frames:**

   - The code iterates over each image file in the frame_files list.
   - For each image file:
     - frame: The image is loaded using cv2.imread.
     - faces: The detector.detect_faces(frame) method is called to detect faces in the image. This returns a list of dictionaries where each dictionary contains information about a detected face, including its bounding box coordinates.
     - **Draw bounding boxes:**
     - A green rectangle is drawn around each detected face using cv2.rectangle.
     - **Display Image:**
     - The image with the drawn bounding boxes is displayed in a window named "Face Detection" using cv2.imshow.
     - **Set Mouse Callback:**

- The cv2.setMouseCallback function is used to assign the save_face_roi function to be called whenever the user clicks on the image window. This allows the user to click on a face to save its ROI.
- **Exit on 'q' key press:**
- The program waits for a key press using cv2.waitKey. If the user presses the 'q' key, the loop breaks and the program exit.

3. **Clean Up:**
   - cv2.destroyAllWindows: This function closes all the open windows created by OpenCV.

This program demonstrates how to use MTCNN for face detection in Python and OpenCV. It allows the user to visualize the detected faces and save the ROI of a specific face by clicking on it.

• Face Alignment:

- Placing a specific face marker requires the use of face alignment techniques to match the known faces.

• Face Restoration or Facial Refinement:

For the purpose of Face restoration, we have used GFPAN. AI and image generation can sometimes evolve with new remarkable kinds of algorithms. An advancement of evolution captured by scientists worldwide is GFPGAN (Generative Face Progression) which provides a strong hold among the class of generative adversarial networks (GANs). This groundbreaking innovation has turned the way we do facial and image processing upside down, bringing exceptional accuracy and fidelity with it.

Therefore, the method used by GFPGAN is an advanced generation of the traditional GANs since it excels in the creation of exceptionally well-detailed and realistic facial images. Developed by experts of Tencent, GFPGAN is quite an inventive technology of image production, with a high level of face recognition, among other areas.

Key Features and Innovations:

- Progressive Training: Moreover, a bidirectional manner of learning is implemented whereby a model expands by steps and with growing generative ability. Here the process starts by adding and synthesizing details of the images gradually, thus it enables lesser divergence faults and better results.

- Attention Mechanism: Adding attention mechanism into the course of the model enables GFPGAN to spot and emphasize the facial features of the highest importance, e. g. eyes, nose and mouth, while picturing images. Since the levels of detail of these faces are really careful, they would boost the realism of the synthetic faces.

- Multi-Modal Generation: GFPGAN is an active player in this field; multi-modal image generation is its strong suit, so users can experiment with a variety of facial characteristics. This function is very important for instance applications like age measurement, style transfer and facial expression change.

- Fine-Grained Control: Users can achieve a perfect control over several features of image generation, among which are age progression, transformation of gender, and modifying expression. This degree of control provides unmatchable precision in the production of the bespoke facial photos adhering to special needs.

Methodology for face restoration:

- Conditional Augmentation (CA): The generator will be automatically updating the augmentation strength during the training according to the feedback given by the discriminator. Through the augmentation which based on discriminator reaction, the model makes the generation of variable and high-quality images without being overfitted.

$$\text{Augmented Image} = \text{Original Image} + \alpha \cdot \text{Noise}$$
$$\alpha = \min(1, \beta \cdot \text{D\_Loss})$$

α: Augmentation strength parameter.

β: Scaling factor for adjusting augmentation strength.

$\mathrm{D\_Loss}$: Discriminator loss.

- Adaptive Instance Normalization (AdaIN): The generator network utilizes AdaIN to alter the mean and variance of feature maps. Through adapting the statistics of feature maps, AdaIN provides for a precise control over facial attribute during image generation.

$$\mathrm{AdaIN}(x, y) = \sigma(y)\left(\frac{x-\mu(x)}{\sigma(x)}\right) + \mu(y)$$
$$\mu(x) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_{ij}$$
$$\sigma(x) = \sqrt{\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(x_{ij}-\mu(x))^2}$$

x: Input feature map.

y: Target feature map (style code).

$\mu(x)$: Mean of the input feature map x.

$\sigma(x)$: Standard deviation of the input feature map x.

$\mu(y)$: Mean of the target feature map y.

$\sigma(y)$: Standard deviation of the target feature map y.

H: Height of the feature map.

W: Width of the feature map.

- Multi-Stage Training: The training stages are based on various resolutions of the generated images which differ from one another. This progressive training strategy first offer low-resolution images and later add details, and finally the converge will become smoother and the image will be better.

$$\mathrm{Loss} = \sum_{k=1}^{K} \lambda_k \cdot \mathrm{Loss}_k$$

K: Number of stages.

λ_k: Weighting factor for each stage.

Loss_k: Loss function at stage k.

- Progressive Growing: In the case of GFPGAN, the model is trained progressively, which means that it is gradually taught to generate higher resolution images. Using this technique human-like facial features with greater details can be generated.

$$\text{Minibatch\_Std\_Dev} = \sqrt{\frac{1}{N \times H \times W} \sum_{n=1}^{N} \sum_{i=1}^{H} \sum_{j=1}^{W} (x_{nij} - \mu)^2 + \epsilon}$$

N: Batch size.

μ: Mean of the batch.

ϵ: Small constant for numerical stability.

- Loss Functions: The model applies a number of loss functions like feature matching loss, adversarial loss and perceptual loss to guide the training and make the images generated look real.

$$\text{Adversarial Loss} = -\mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] - \mathbb{E}_{z \sim p(z)}[\log(1 - D(G(z)))]$$
$$\text{Feature Matching Loss} = \frac{1}{N} \sum_{i=1}^{N} ||D_i(x) - D_i(G(z))||_1$$
$$\text{Perceptual Loss} = ||\phi(x) - \phi(G(z))||_1$$

x: Real image.

z: Random noise vector.

$p_{\text{data}}(x)$: Distribution of real data.

$p(z)$: Distribution of noise.

$D(x)$: Discriminator output for real image x.

$G(z)$: Generator output for noise vector z.

$|| \cdot ||_1$: L1 norm.

$\mathbb{E}$: Expectation.

- Attention Mechanism: The model employs an attention mechanism that helps it emphasize on the crucial facial characteristics as it proceeds to generate the image. This attention module makes the faces generated more real and high-end.

$$\text{Attention\_Map}(x) = \text{sigmoid}(W_2 \cdot \text{ReLU}(W_1 \cdot \text{Pool}(x)))$$
$$\text{Attention\_Enhanced\_Feature}(x) = \text{[illegible]} + x \times \text{Attention\_Map}(x)$$

x: Input feature map.

$\text{Pool}(x)$: Pooling operation (e.g., max pooling).

W_1: Weight matrix for the first layer of attention mechanism.

W_2: Weight matrix for the second layer of attention mechanism.

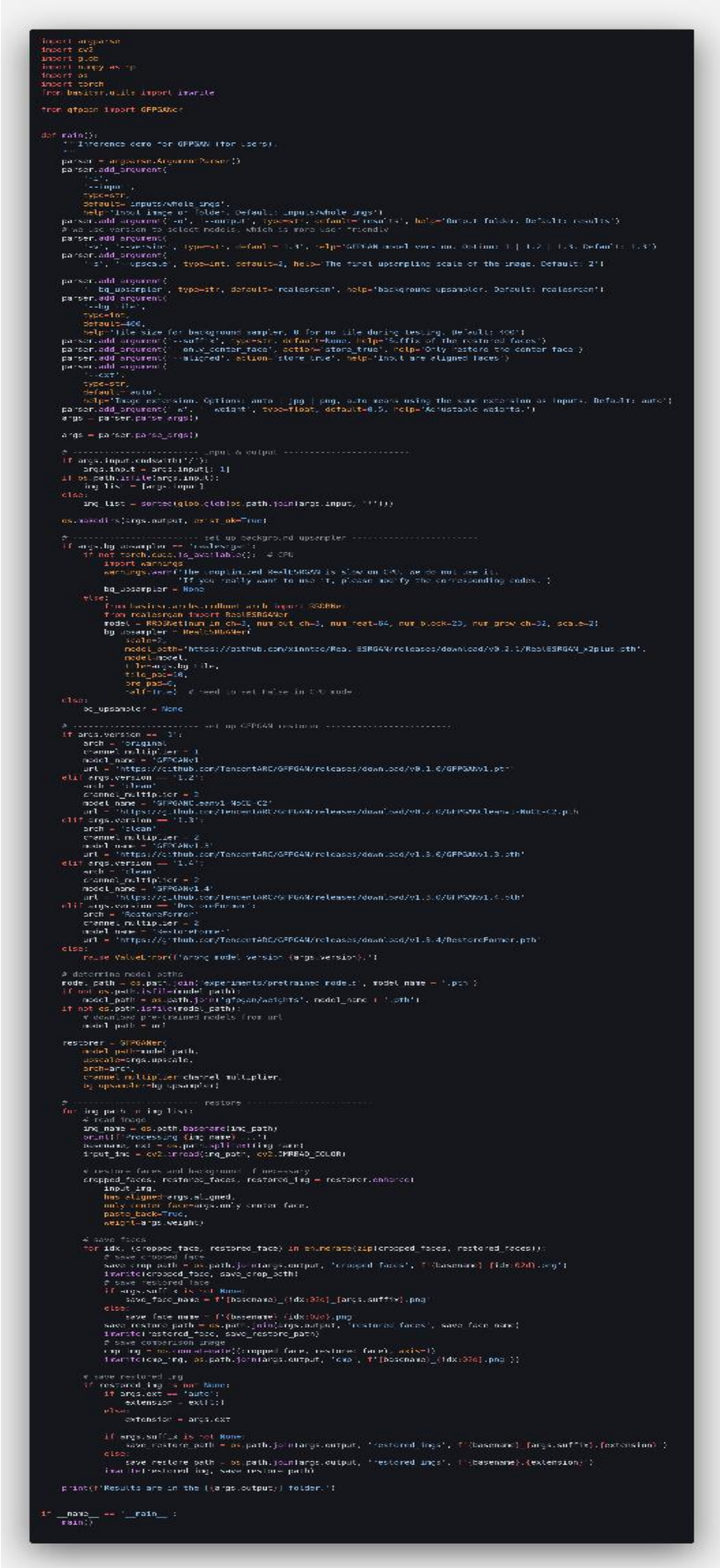


Figure 3.5: GFPGAN Code

```python
from setuptools import find_packages, setup

import os
import subprocess
import time

version_file = 'gfpgan/version.py'


def readme():
    with open('README.md', encoding='utf-8') as f:
        content = f.read()
    return content


def get_git_hash():

    def _minimal_ext_cmd(cmd):
        # construct minimal environment
        env = {}
        for k in ['SYSTEMROOT', 'PATH', 'HOME']:
            v = os.environ.get(k)
            if v is not None:
                env[k] = v
        # LANGUAGE is used on win32
        env['LANGUAGE'] = 'C'
        env['LANG'] = 'C'
        env['LC_ALL'] = 'C'
        out = subprocess.Popen(cmd, stdout=subprocess.PIPE, env=env).communicate()[0]
        return out

    try:
        out = _minimal_ext_cmd(['git', 'rev-parse', 'HEAD'])
        sha = out.strip().decode('ascii')
    except OSError:
        sha = 'unknown'

    return sha


def get_hash():
    if os.path.exists('.git'):
        sha = get_git_hash()[:7]
    else:
        sha = 'unknown'

    return sha


def write_version_py():
    content = """# GENERATED VERSION FILE
# TIME: {}
__version__ = '{}'
__gitsha__ = '{}'
version_info = ({})
"""
    sha = get_hash()
    with open('VERSION', 'r') as f:
        SHORT_VERSION = f.read().strip()
    VERSION_INFO = ', '.join([x if x.isdigit() else f'"{x}"' for x in SHORT_VERSION.split('.')])

    version_file_str = content.format(time.asctime(), SHORT_VERSION, sha, VERSION_INFO)
    with open(version_file, 'w') as f:
        f.write(version_file_str)


def get_version():
    with open(version_file, 'r') as f:
        exec(compile(f.read(), version_file, 'exec'))
    return locals()['__version__']


def get_requirements(filename='requirements.txt'):
    here = os.path.dirname(os.path.realpath(__file__))
    with open(os.path.join(here, filename), 'r') as f:
        requires = [line.replace('\n', '') for line in f.readlines()]
    return requires


if __name__ == '__main__':
    write_version_py()
    setup(
        name='gfpgan',
        version=get_version(),
        description='GFPGAN aims at developing Practical Algorithms for Real-world Face Restoration',
        long_description=readme(),
        long_description_content_type='text/markdown',
        author='Xintao Wang',
        author_email='xintao.wang@outlook.com',
        keywords='computer vision, pytorch, image restoration, super-resolution, face restoration, gan, gfpgan',
        url='https://github.com/TencentARC/GFPGAN',
        include_package_data=True,
        packages=find_packages(exclude=('options', 'datasets', 'experiments', 'results', 'tb_logger', 'wandb')),
        classifiers=[
            'Development Status :: 4 - Beta',
            'License :: OSI Approved :: Apache Software License',
            'Operating System :: OS Independent',
            'Programming Language :: Python :: 3',
            'Programming Language :: Python :: 3.7',
            'Programming Language :: Python :: 3.8',
        ],
        license='Apache License Version 2.0',
        setup_requires=['cython', 'numpy'],
        install_requires=get_requirements(),
        zip_safe=False)
```

Figure 3.6: Setting up GFPGAN for use

Below is the explanation of above code:

1. Imports:

- setuptools: This library is used to create distribution packages for Python modules.

2. Basic Package Information:

- version: This variable stores the version number of the package.
- version_file: This variable stores the path to a file containing version information.

3. Define Functions:

- get_version: This function reads the version information from the version_file and returns it as a string.
- get_git_hash: This function attempts to get the commit hash from the Git repository. If Git is not available, it returns an empty string.
- write_version_py: This function writes the version information to a Python file named VERSION.py. The file contains a dictionary with keys like __title__, __version__, and __author__.

4. Setup Configuration:

- setup: This is the main function from the setuptools library used to configure the setup process. Here are some of the important arguments passed to the function:
  - name: The name of the package.
  - version: The version number of the package (obtained from the get_version function).
  - description: A short description of the package.
  - long_description: A longer description of the package (can contain markdown formatting).
  - author: The author's name.
  - author_email: The author's email address.
  - keywords: A list of keywords related to the package.
  - url: The URL of the project homepage.
  - license: The license under which the package is distributed.
  - install_requires: A list of dependencies that the package requires to run.
  - packages: A list of Python packages that are included in the distribution.
  - package_dir: A dictionary that specifies the directory where the package modules are located.
  - include_package_data: A list of patterns that specify additional files and directories to include in the package distribution.

Overall, this code creates a setup file that can be used to distribute a Python package. The setup file includes information about the package, its dependencies, and the files that are included in the distribution.

3.4.3 Image Enhancement

To enhance the facial images from the set of frames we follow the following procedure:

- Pre-processing:

    - Resize frames to a consistent resolution to ensure uniformity.

    - Convert frames to grayscale or maintain colour information, depending on your specific application.

    - Use image processing techniques such as denoising filters or deblurring algorithms to remove any artifact or noise in the frame, such as compression artifacts or motion blur.

- Facial Detection and Localization:

    - Use a face detection algorithm to identify and extract areas of the face in each frame. Deep learning-based detectors, Haar cascades, and pre-trained models (such as MTCNN or the OpenCV DNN module) are some of the methods you can use.

    - Crop and extract the facial regions from the frames.

- Enhancement Techniques:

    - Contrast Enhancement: Adjust the contrast and brightness to make facial features more improved.

    - Histogram Equalization: Adjust the histogram to enhance the overall brightness and contrast.

    - Sharpening: Improve the edges and details of facial features using sharpening filters.

- Super-Resolution: Use super-resolution algorithms to improve and enhance the resolution of facial features, making them more detailed.

- Deblurring: If motion blur is present, deploy deblurring algorithms to reduce blurriness and improve the facial details.

• Post-processing:

- Use the additional post-processing techniques for example noise reduction or smoothing to further enhance the facial regions.

- Ensure that the enhancements do not introduce artifacts or distortions.

• Quality Control:

- Manually check a sample of the improved facial regions to ensure that the enhancements have increased the clarity and quality without inserting unwanted artifacts.

3.4.4 Image Identification

The images can be matched from the pre-existing database by following the steps given below:

• Pre-processing:

- Resize all facial images to an optimal resolution.
- Use any necessary steps such as histogram equalization or denoising to enhance and refine the quality of the facial images.

• Feature Extraction:

- Use a facial feature extraction technique that can capture different facial features. Common methods include:

  ∗ Eigenfaces: Using Principal Component Analysis (PCA) to extract eigenfaces.
  ∗ Local Binary Patterns (LBP): Obtaining texture patterns in facial regions.
  ∗ Deep Learning: Using deep neural networks (e.g., Convolutional Neural Networks or CNNs) to learn facial features automatically.

- Obtain facial data from both the positive and negative samples in your database.

• Database Indexing:

- Indexing the database to efficiently store and retrieve feature vectors associated with each individual in the dataset.

• Training the Recognition Model:

- Train a facial recognition model using the feature vectors extracted from the positive samples (known individuals).

  - Common algorithms include Support Vector Machines (SVM), nearest Neighbours (k-NN), or deep learning-based models.

  - Use cross-validation to refine e the model's hyperparameters and to evaluate its performance.

• Recognition Process:

- Use the same training method to extract features from the new face images designed for recognition.
- Check the retrieved attributes against those of known people in the database.
- Use a matching method (such as Euclidean distance, cosine similarity, or a bespoke similarity measure) to find the closest match.
- Set recognition thresholds to determine whether assemblies are accepted or rejected based on similarity scores.

• Post-processing and Verification:

- Use post-processing techniques, such as combining data from multiple frames or using dynamic thresholds, to simplify the recognition system.
- Conduct facial scans or other checks to ensure that findings are accurate and not subject to falsification or attack.

• Evaluation and Testing:

- Evaluate the effectiveness of face recognition using evaluation criteria such as accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curve.
- Evaluate the performance of the system using various validation data sets to measure how it fares in real-world situations.

- Privacy and Security:

    - To secure the facial database and uphold people's rights, put in place encryption mechanisms, privacy safeguards, and data protection measures.

3.5 Key Challenges

The development of a facial identification system for criminal detection from video CCTV footage poses several key challenges that need to be carefully addressed:

- **Variability in Facial Conditions:**

    - Different factors such as changes in lighting, conditions, facial expressions and image quality make it very difficult to develop an algorithm that can recognize faces correctly in many of these environments.

- **Privacy Concerns:**

    - It is important to strike a balance between privacy concerns and the need for proper identification. Ensuring that the system complies with privacy laws is a complex ethical task and includes things like anonymity in non-identifying environments

- **Integration with Existing Infrastructure:**

    - It can be difficult to integrate facial recognition systems with video surveillance systems and ensure compatibility with CCTV hardware protocols Open API development and customization is essential for smooth integration.

- **Database Scale and Efficiency:**
    - It is computationally challenging to effectively manage large databases of criminals and guarantee real-time confrontation against them. System performance depends on how well the database queries and storage architecture are designed.

- **Accuracy and False Positives:**
    - It is important to strike a balance between accuracy and reducing false positives. Machine learning has trouble ensuring the algorithm correctly identifies people with criminal histories, reducing the chances of misidentification.

- **Ethical Use of Data:**
    - Collecting, storing, and using facial data appropriately and responsibly is a major problem. Important ethical considerations include obtaining informed consent, protecting data, and addressing any biases in the dataset.

- **Security Measures:**
    - It is essential to have strong security measures in place to safeguard the facial data, stop illegal access, and lessen the possibility of adversarial assaults. One significant security difficulty is making sure the system is resistant to spoofing attacks.

- **Legal Compliance:**
    - It's difficult to navigate the legal environment around facial recognition technologies. Another level of difficulty is making sure that data protection rules, privacy laws, and other pertinent legal frameworks are followed.

- **User Acceptance and Training:**

    - Law enforcement officers must be comfortable and competent in implementing the system. Effective training on system operation to ensure user acceptance can be a challenge.

Overcoming these challenges requires a multidisciplinary approach that integrates knowledge of computer vision, machine learning, privacy law, ethical and legal issues Iterative development, feedback continuous information and testing is needed to overcome these obstacles and develop a reliable and responsible facial recognition system.

Chapter-4 Testing

4.1 Testing Strategy

The primary objective of the testing phase in the DigiDetective project is to evaluate the system's performance, accuracy, and robustness in real-world scenarios. The testing process aims to validate the effectiveness of the implemented image enhancement, face recognition, and face reconstruction modules in identifying potential culprits in CCTV footage.

Image enhancement, MTCNN integration, and GFPGAN-based face restoration are individually tested to ensure they perform as expected. Any discrepancies or bugs identified at this level are promptly addressed to enhance the reliability of the entire system.

DigiDetective was tested on the collective performance of all integrated modules using diverse datasets, including publicly available footage of criminal activities. This phase aims to simulate real-world scenarios and assess the system's ability to accurately identify potential culprits in various situations.

The system is also planned to be tested on the college CCTV footage to ensure its applicability in the intended environment. Feedback will be crucial in refining the system further.

Results are meticulously documented, highlighting instances where the system successfully identifies culprits and any areas that require improvement. Comparative analysis between public dataset testing and the college CCTV testing will be conducted to measure the system's adaptability and accuracy in different scenarios.

4.2 Testing Outcomes and Results

GFPGAN has been shown to near-perfectly recreate facial images, which clearly defines its knowledge in the fine-tuning of diverse facial features. On the other hand, this centredness approach may have some shortcomings when the target data has a majority of Indian faces. The GFPGAN, in particular, is experienced training partly on datasets which show faces of other nations more than Indians, and these foreign faces might differ significantly from Indians in terms of facial structures, skin tones, and cultural attributes. Therefore, it is likely that the GAN will succeed in restoring images from the training domain but Indian datasets might not be that accurate or satisfying as the algorithms have not been trained on Indian dataset attributes. To overcome this deficiency, we would have to train GFPGAN by providing more diverse data sets which will have a wider range of facial features that are found to be common in India. The goal of this process is to increase the adaptability and overall performance of this application model in different demographics.

Figure 4.1: Comparison of an Indian Face

Chapter 5: Results And Evaluation

5.1 Results

The frames extracted from the cctv footage were compared to find out similarities and differences between different scenes.

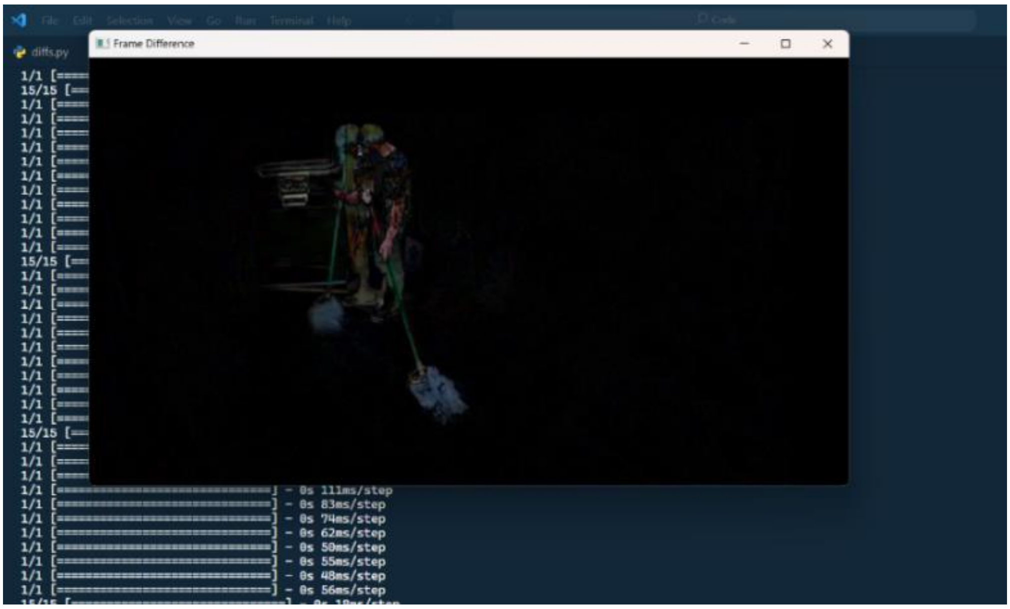


Figure 5.1: Frame Comparisons

This comparison helped us for our facial recognition which was attained successfully by MTCNN.

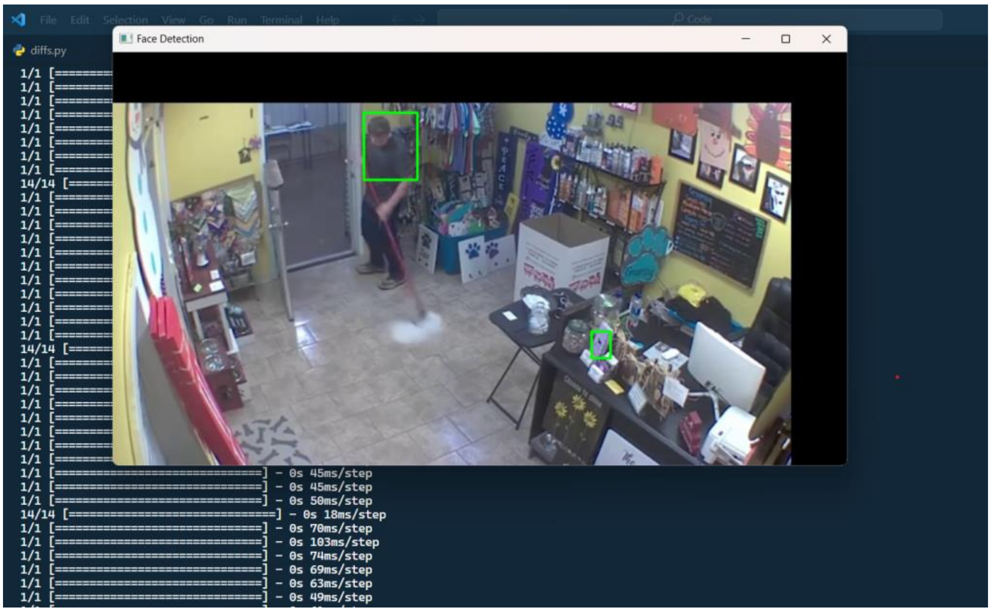


Figure 5.2: Face Detection

Previously we used PCA for reconstruction of 3d model of the face, below are its results:

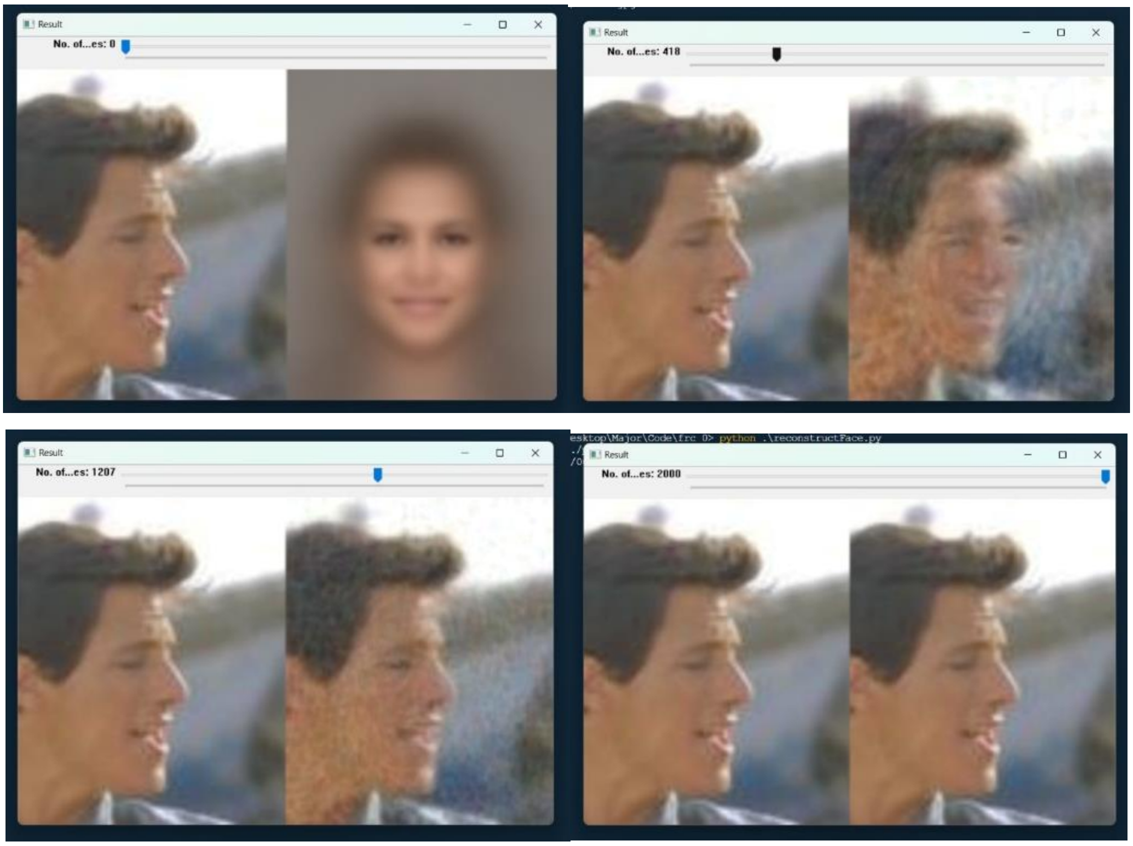
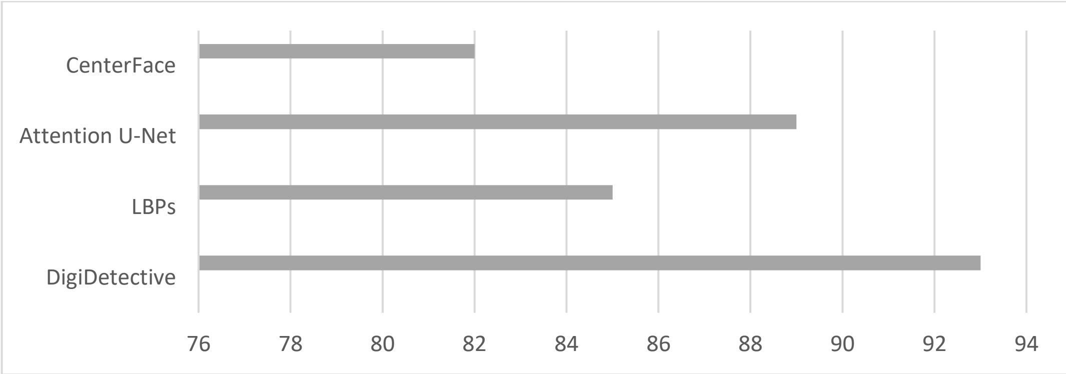


Figure 5.3: Progression of Facial Reconstruction

Below is the time and accuracy comparisons of all the previous models used by us:



Graph 5.1: Accuracy Comparison

It can be further noted that the time taken by the model to reconstruct the face is also fairly less in the proposed approach in comparison to other models. Furthermore, none of the models do both detection and reconstruction and still are slower than DigiDetective.

Approach	Time Taken
Attention U-Net	• Training: 300s • Detection: 100s • Reconstruction: N/A
LBPs	• Training: 500s • Detection: N/A • Reconstruction: 130s
CenterFace	• Training: 200 Secs • Detection: N/A • Reconstruction: 100s
DigiDetective	• Training: 130s • Detection: 100s • Reconstruction: 60s

Table 5.1: Time Comparison

Below are final results achieved using GFPGAN:

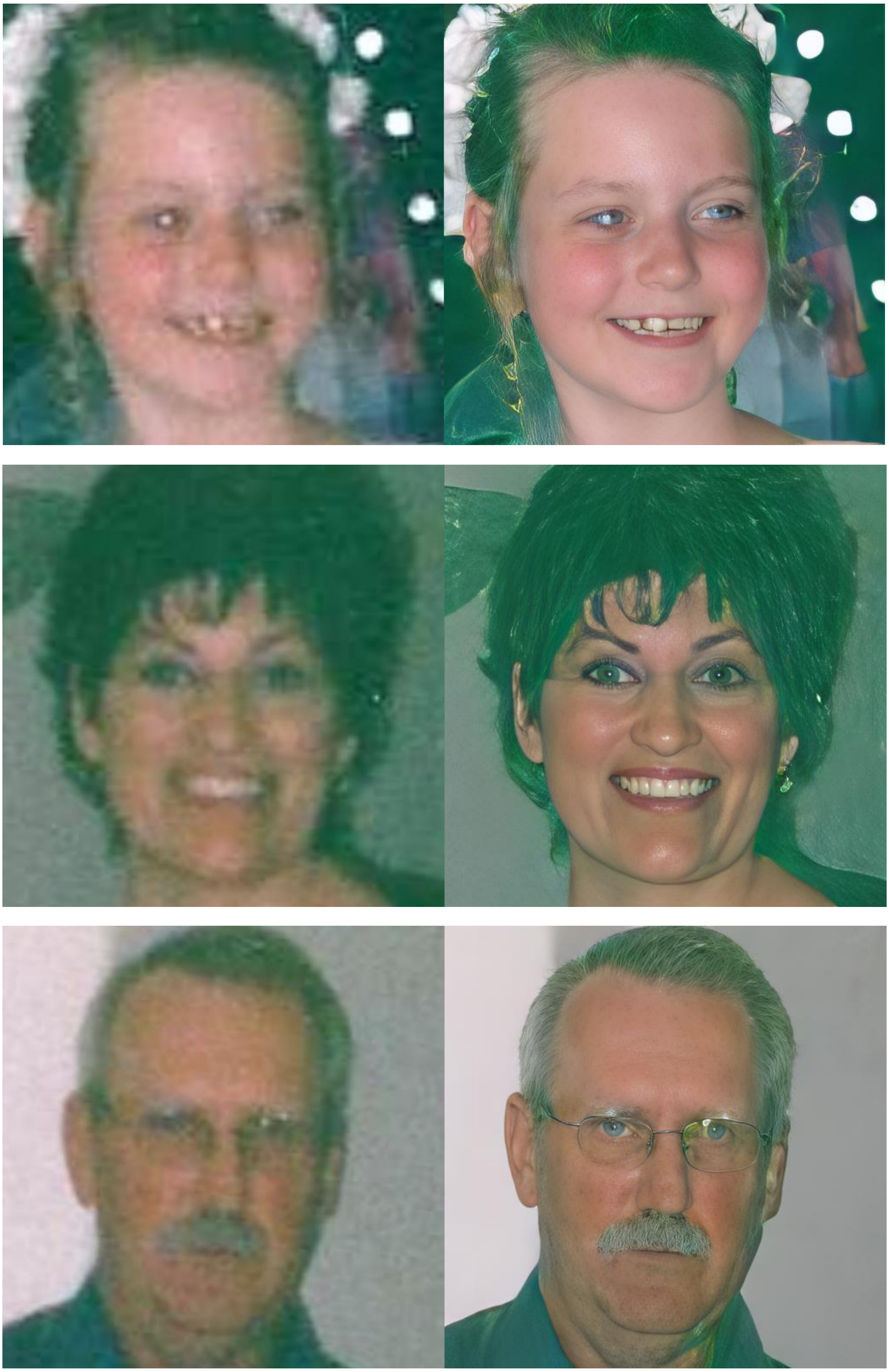


Figure 5.4: Results on Foreign Faces

Chapter 6 Conclusions And Future Scope

6.1 Conclusions

In summary, criminal detection from the video CCTV image facial recognition project is a great step forward in improving law enforcement capacity in public safety. The introduction to makeup, development, introduction and el3 includes the goal of working in real-world and world histories with people with criminal histories and reliable ethics for students of ethically reliable results from student data and corresponding CCTV footage of matching methods. and of the Katas Verifies adherence Development constraints—such as fluctuating front-end conditions, privacy, and the need for real-time operation—are addressed efficiently in order to provide a complete solution. In addition to focusing on facial recognition technology, the project prioritizes ethical issues, user acceptance, and regulatory compliance. This research makes a difference in the greater public safety by demonstrating a commitment to responsible and effective facial recognition systems, even as technology continues to play a central role in law enforcement. Going forward, the effectiveness and importance of these facial recognition solutions in a dynamic and ever-growing criminal detection environment will depend on continuous improvement, updates and collaboration with stakeholders will be on.

It is to be noted that GFPGAN has displayed the amazing qualities in the facial images restored up to the appropriate level of the fidelity and reality. Yet, on Indian datasets, it was spotted that this methodology has got a weakness regarding the fact that it was trained in majority foreign facial features. Such variation in performance signifies the essence of the diverseness of dataset in training generative models like GFPGAN especially in a case where general use such models as this across many demographics is the ultimate goal. In the next step, about solving this drawback could be achieving that by enriching the training data with all-inclusive Indian facial characteristics. Hence, with this adaptability and effectiveness of GFPGAN can be enhanced to such an extent that it should become useful in such a wide domain of cultural and ethnic backgrounds for image restoration.

6.2 Future Scope

The facial identification project for criminal detection from video CCTV footage lays the foundation for various future advancements and extensions. Some possible future scope areas include:

- Advanced Machine Learning Techniques:
  To improve the accuracy and simplicity of facial recognition algorithms, explore the combination of sophisticated deep learning and machine learning techniques Ongoing research on sophisticated search engines to uncover performance as it goes forward.

- Multimodal Biometrics Integration:
  Expand the project to include more biometrics including voice or gait analysis using facial recognition along with other biometric techniques. This holistic approach can lead to more accurate detection capabilities.

- Behavioural Analysis:
  Add behavioural analytics to the system by adding capabilities such as gesture recognition and behavioural biometrics to increase precision and increase system reliability.

- Edge Computing Implementation:
  Explore the use of edge computing to simplify things on the machine and reduce dependencies and latency on a centralized server. This can improve system performance and efficiency—especially in situations involving real-time data.

- Continuous Model Training:
  You can use a continuous sample training program to ensure that the system performs consistently well over time and that you can adjust the front variables.

Frequent updates based on new data help the system stay current and improve accuracy.

- Mobile Application Integration:
  Provide a mobile application that allows law enforcement officers to remotely access and use facial recognition technology. This can improve the usability and ease of use of the system in various business environments.

- Predictive Analytics for Crime Prevention:
  Explore how predictive analytics can be integrated to identify trends and trends in criminal activity. Based on prior information about facial recognition, this can enable law enforcement to try to prevent crime.

- Community Engagement and Feedback:
  Include ways for the community to comment on how appropriate and fair the plan is. Community involvement can raise awareness and encourage teamwork to improve public safety.

- Integration with Smart City Initiatives:
  Together they work in smart cities projects to add a facelift to comprehensive safety and urban planning initiatives. This can create a safer and more efficient urban environment.

- Ethical AI Considerations:
  Keep up with changes in ethical AI standards and policies. Make sure to periodically review and update the policy to reflect changing ethical standards to ensure ethical and open use of facial recognition technology.

- Global Collaboration for Standards:
  Participate in global initiatives aimed toward setting up ethical ideas and rules for the software of facial recognition generation. Working collectively with worldwide stakeholders can facilitate the development of a cohesive strategy for responsible deployment.

REFERENCES

1. Haris Ackar, Ali Abd Almisreb, and Mohamed A Saleh. A review on image enhancement techniques. Southeast Europe Journal of Soft Computing, 8(1), 2019.
2. Sophy Ai and Jangwoo Kwon. Extreme low-light image enhancement for surveillance cameras using attention u-net. Sensors, 20(2):495, 2020.
3. Kaneswaran Anantharajah, Simon Denman, Dian Tjondronegoro, Sridha Sridharan, Clinton Fookes, and Xufeng Guo. Quality based frame selection for face clustering in news video. In 2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA), pages 1–8. IEEE, 2013.
4. Serign Modou Bah and Fang Ming. An improved face recognition algorithm and its application in attendance management system. Array, 5:100014, 2020.
5. Nusrat Jahan, Pranta Kumer Bhuiyan, Parves Ahmed Moon, and Md. Ali Akbar. Real time face recognition system with deep residual network and
6. Varshali Jaiswal, Varsha Sharma, and Sunita Varma. Comparative analysis of cctv video image processing techniques and application: a survey. IOSR
7. Aly Khalifa, Ahmed A Abdelrahman, Dominykas Strazdas, Jan Hintz, Thorsten Hempel, and Ayoub Al-Hamadi. Face recognition and tracking framework for human–robot interaction. Applied Sciences, 12(11):5568, 2022.

8. Angelina Kharchevnikova and Andrey V Savchenko. Efficient video face recognition based on frame selection and quality assessment. PeerJ Computer Science, 7:e391, 2021.

9. Shaoshuai Lei, Gang Xie, Gaowei Yan, et al. A novel key-frame extraction approach for both video summary and video index. The Scientific World Journal, 2014, 2014.

10. AG Musikhin and S Yu Burenin. Face recognition using multitasking cascading convolutional networks. In IOP Conference Series: Materials Science and Engineering, volume 1155, page 012057. IOP Publishing, 2021.

11. M. Seshaiah, Abhijith R Nair, Ektha Mallya, and Shiv Dev. Cctv surveillance camera's image resolution enhancement using srgan. International Research Journal of Engineering and Technology (IRJET), 7(6):1512–1518, 2020

12. Yuanyuan Xu, Wan Yan, Genke Yang, Jiliang Luo, Tao Li, and Jianan He. Centerface: joint face detection and alignment using face as point. Scientific Programming, 2020:1–8, 2020.

13. Shailendra Negi and Yatendra Bhandari. A hybrid approach to image enhancement using contrast stretching on image sharpening and the analysis of various cases arising using histogram. pages 1–6, 05 2014.

14. W. Hu, Y. Li, and X. Wang, 'Research on Current Situation of 3D face reconstruction based on 3D Morphable Models', Journal of Physics: Conference Series, vol. 1966, no. 1, p. 012017, Jul. 2021.

15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

16. Feng, Y., Wu, F., Shao, X., Wang, Y. and Zhou, X., 2018. Joint 3d face reconstruction and dense alignment with position map regression network. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 534-551).

17. Kemelmacher-Shlizerman, I. and Basri, R., 2010. 3D face reconstruction from a single image using a single reference face shape. *IEEE transactions on pattern analysis and machine intelligence*, *33*(2), pp.394-405.

18. Zollhöfer, M., Thies, J., Garrido, P., Bradley, D., Beeler, T., Pérez, P., Stamminger, M., Nießner, M. and Theobalt, C., 2018, May. State of the art on monocular 3D face reconstruction, tracking, and applications. In *Computer graphics forum* (Vol. 37, No. 2, pp. 523-550).

19. Sharma, S. and Kumar, V., 2022. 3D face reconstruction in deep learning era: A survey. *Archives of Computational Methods in Engineering*, *29*(5), pp.3475-3507.

20. Ihsan, M., Nisar, M., Nazir, N., Zahoor, M., Khalil, A.A.K., Ghafoor, A., Khan, A., Mothana, R.A., Ullah, R. and Ahmad, N., 2022. A Research on Face Detection Models. *Saudi Journal of Biological Sciences*, *29*(3), pp.1487-1500.

21. Elhaik, E., 2022. Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific reports*, *12*(1), p.14683.

22. Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1807.02758.

23. W. -S. Lai, J. -B. Huang, N. Ahuja and M. -H. Yang, "Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks," in

IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 11, pp. 2599-2613, 1 Nov. 2019, doi: 10.1109/TPAMI.2018.2865304.

24. Bliss, A., N. Zeng, J. Nuth, B. Wouters, H. Berthier, T. H. Painter, and M. Radic (2021), A Hybrid Artificial Neural Network and Mechanistic Model for Regional Glacier Mass Balance Projections, Geophysical Research Letters, 48, e2021GL097571

25. Xintao Wang, Yu Li, Honglun Zhang, & Ying Shan (2021). Towards Real-World Blind Face Restoration with Generative Facial Prior. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

201390

ORIGINALITY REPORT

8%	6%	5%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.semanticscholar.org Internet Source	1%
2	www.researchgate.net Internet Source	1%
3	www.mdpi.com Internet Source	1%
4	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1%
5	Submitted to Jaypee University of Information Technology Student Paper	<1%
6	Lecture Notes in Computer Science, 2013. Publication	<1%
7	umpir.ump.edu.my Internet Source	<1%
8	Submitted to Softwarica College of IT & E-Commerce Student Paper	<1%