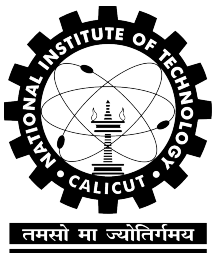
Suspect Detection In Crowd

CS4099 Project Final Report

Submitted by

Shreyas S N	**Reg No: B200773CS**
Vinit Kumawat	**Reg No: B200819CS**
Mohammad Shaheem C	**Reg No: B200721CS**

Under the Guidance of
Dr. Santosh Kumar Behera

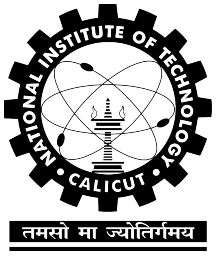


Department of Computer Science and Engineering
National Institute of Technology Calicut
Calicut, Kerala, India - 673 601

May 8, 2024

NATIONAL INSTITUTE OF TECHNOLOGY CALICUT, KERALA, INDIA - 673 601

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



2024

CERTIFICATE

Certified that this is a bonafide record of the project work titled

SUSPECT DETECTION IN CROWD

done by

Shreyas S N

Vinit Kumawat

Mohammad Shaheem C

centring of eighth semester B. Tech in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering of the National Institute of Technology Calicut

Project Guide

Dr. Santosh Kumar Behera

Assistant Professor

05/05/2024

DECLARATION

I hereby declare that the project titled, **Suspect Detection In Crowd**, is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or any other institute of higher learning, except where due acknowledgement and reference has been made in the text.

Place : NIT Calicut
Date : May 8, 2024

Signature : 
Name: Shreyas S N
Reg. No. B200773CS

Signature : 
Name : Vinit Kumawat
Reg. No. B200819CS

Signature : 
Name: Mohammad Shaheem C
Reg. No. B200721CS

Abstract

Along with the fast-growing economy of a developing country as far as India is concerned, an emerging number of crimes and criminal offences have been reported. Regrettably, in numerous instances, suspects manage to evade justice due to insufficient testimonials and a lack of timely communication channels. Using deep learning techniques such as Convolutional Neural Networks (CNNs) in face recognition has shown remarkable performance in accurately identifying and verifying faces from images and videos. The advancements have made it possible to verify faces under challenging situations like variations in lighting, poses, facial expressions, and the presence of accessories. Our project aims to prevent crime by identifying suspects in crowded locations with a suspect database and video surveillance system in real time. With this innovative approach, we create a deep learning model that has good robustness against occlusion and low resolution in face detection, effectively expands the distance between classes in face recognition, and improves recognition accuracy.

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to Dr Santosh Kumar Behera for his unwavering support and invaluable guidance during the culmination of our final-year project. His expertise and encouragement have profoundly shaped its development, and we are sincerely appreciative of his dedication. Additionally, we extend our gratitude to friends for their steadfast support throughout this journey, as their encouragement has been a constant source of strength. We are also thankful to the esteemed faculty members whose insightful input has significantly enriched the quality of our work. Together, their collective contributions have not only enhanced the calibre of my project but have also fostered a profound learning experience. We are truly fortunate to have been surrounded by such a supportive and knowledgeable community, and we are deeply grateful for their contributions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In today's increasingly security-conscious society, the demand for effective suspect detection measures in crowded public spaces has never been greater. Deep learning, a subset of artificial intelligence (AI), emerges as a promising avenue for bolstering the accuracy and efficiency of suspect identification. This research endeavours to explore the seamless integration of deep learning techniques into existing systems, with a particular focus on enhancing suspect identification through facial recognition and witness-based methodologies.

The pivotal role of deep learning methods, such as MTCNN for detection and FaceNet for embeddings, cannot be overstated in the quest for precise and unobtrusive criminal identification. Leveraging these sophisticated algorithms, we aim to push the boundaries of accuracy and reliability in suspect identification, ensuring that security measures remain effective and non-invasive.

Beyond individual suspect identification, our research also delves into crowd attention analysis, harnessing the power of deep learning to develop intelligent systems capable of real-time monitoring. By employing advanced algorithms, we seek to create a comprehensive solution that not only identifies suspects but also enhances situational awareness in crowded environments.

Crucially, our research addresses the pressing need for an efficient sus-

pect detection system that can operate seamlessly across diverse conditions. Whether dealing with dynamic datasets or imperfect images, our goal is to develop a robust framework that transcends the limitations of previous approaches. Through meticulous analysis and synthesis of existing literature, we aim to provide a comprehensive overview of the state-of-the-art techniques and identify areas where further advancements can be made.

In essence, our research endeavours to bridge the gaps in existing knowledge and technology, paving the way for the development of highly effective suspect detection systems capable of operating reliably under the most challenging conditions.

Chapter 2

Literature Survey

Multiple research papers have prioritized tackling the complexities of face detection and recognition in densely populated environments. Below are summarized synopses of research papers that closely align with our intended work.

This research paper [1] uses advanced machine learning for real-time crime detection and identification, driven by the idea that criminals may exhibit specific facial traits. The study employs a comparative analysis of deep learning models, including VGG-16, VGG-19, and InceptionV3, particularly emphasizing male images. The VGG CNN models achieve a remarkable 99.5 percent accuracy in identifying criminal faces, and the approach benefits from pre-trained models, enhancing accuracy without extensive training. While the research shows promise in real-time crime detection, it has limitations. It is restricted to a limited dataset size and may result in misclassifications. To enhance the approach, a larger and more diverse dataset is needed to improve the accuracy of multiple deep-learning algorithms for criminal detection and prevention.

The paper [2] introduces a real-time face recognition system for criminal identification, achieving remarkable results with high accuracy. It utilizes machine learning and deep neural networks, specifically MTCNN for facial

landmark detection and FaceNet for embedding facial features, outperforming conventional methods. The work of this paper has limitations in real-time dynamic dataset handling, and it requires enhancements to identify multiple faces from blurry or cropped images efficiently.

The paper [3] presents an intelligent crowd attention detection system using face detection technology. This system employs Haar-like features and the Adaboost algorithm to detect faces, providing a mathematical expected value for crowd attention. In an experiment, it effectively monitored crowd attention, revealing that video engagement increased over time. However, this system needs improvements in handling varying attention levels in different contexts. Additionally, it could benefit from further exploration of Big Data technology for long-term crowd attention trends and more accurate predictions of crowd behaviour, promoting its application in various sectors.

The paper [4] introduces attribute-based face recognition, showing promise for law enforcement and witness identification. Using 46 facial attributes and automated extraction, it achieves accuracy comparable to sketch recognition, providing an effective method for suspect identification. The work of this paper [4] faces constraints in its application to real-world operational scenarios and may encounter challenges in handling low-quality imagery. Furthermore, there's room for improvement in dealing with incomplete attribute sets and implementing confidence-based matching for enhanced identification accuracy.

[5] The authors of this paper introduced the Deep Cascade Model (DCM) for face recognition to address limitations seen in existing methods such as Sparse Representation Classification (SRC), Nearest Mean Residue (NMR), and Deep Learning (DL). SRC and NMR were noted for not fully leveraging coding error vector information, while DL's reliance on extensive data and computational resources made it less suitable for small-scale data. The DCM merges hierarchical learning, nonlinear transformation, and multi-layer structure of DL with discriminative feature abstraction of SRC and NMR,

resulting in enhanced face recognition performance, particularly for scenarios with limited data. The DCM's application is geared towards robust face recognition with feature extraction under challenging conditions, especially for small-scale datasets. The model adeptly employs multi-level image coding for feature representation and integrates existing representation methods to improve overall performance. Its versatility allows for effective deployment across a spectrum of demanding scenarios, showcasing its superiority over state-of-the-art models through efficient utilization of effective pooling functions, a hierarchical SoftMax vector coding module, and a versatile Getting New Feature (GNF) operator.

[6] This study presents a thorough review and evaluation of different techniques. The focus is on both accuracy and computational efficiency, introducing a novel metric to gauge the computational cost-effectively. The aim is to provide valuable insights for selecting appropriate face detection methods and guiding future developments in this domain, specifically targeting applications like face recognition, face tracking, and facial expression recognition. The experimental results and metrics comparative analysis showcase the superiority of deep learning in face detection, with the MTCNN demonstrating remarkable performance. Furthermore, the paper emphasizes how modern deep learning models effectively leverage extensive face datasets, achieving performance levels that rival or even surpass human face recognition capabilities.

[7] The proposed approach aims to improve face detection and recognition by addressing challenges such as mutual face occlusion, crowd scenarios, handling low-resolution images, and accommodating varying face proportions due to camera distances. Employing a Multi-Task Cascaded Convolutional Network (MTCNN) facilitates robust face detection. Integration of Soft Non-Maximum Suppression (Soft-NMS) enhances detection accuracy, particularly in occlusion scenarios. Super-resolution network is employed to improve feature representation for low-resolution face images. Experimental

results demonstrate the effectiveness of the approach, showcasing its ability to handle occlusion, enhance recognition accuracy, and robustly process low-resolution images. Evaluation of established datasets, including WIDER FACE, LFW, and Yale face databases, highlights its superior performance when compared to state-of-the-art methods, affirming its potential to advance face detection and recognition in practical settings.

[8] The paper addresses challenges faced during large-scale Unconstrained face recognition. Labeling huge amounts of data for feeding supervised deep learning algorithms is expensive and time-consuming. Real-world face recognition datasets often have unbalanced pose distributions, making it difficult to improve recognition performance. Several research attempts have been made to employ synthetic profile face images as augmented extra data to balance the pose variations. However, learning directly from synthetic images can be problematic as synthetic data often lacks realism with texture loss and artefacts. For this face recognition system, the simulator extracts face RoI (Region of Interest), performs saliency prediction (i.e., face/background segmentation), localizes 68 landmark points, and produces synthetic faces with arbitrary poses, which are fed to DA-GAN for realism refinement. DA-GAN uses a fully convolutional skip-net (modified to an FCN) as the generator and an autoencoder as the discriminator. The dual agents are responsible for discriminating real versus fake (minimizing the adversarial loss function) and preserving identity information (minimizing the identity perception loss function).

[9] When applying face recognition models in real-world scenarios, there are still many challenges like extreme illumination, rare head pose, low resolutions, and occlusions. This paper adopts a "large margin cosine loss" as its training face recognition loss function which is an improvement over the traditional SoftMax function. The state-of-the-art model for occluded face recognition (PDSN) needs K2 deep face models to be trained separately at the training stage to learn the dictionary, which makes it inefficient and

time-consuming for training. Introduces a FROM (Face Recognition with Occlusion Mask) approach as a single-network solution, which can be trained end to end, for occluded face recognition. The model first takes a mini-batch which consists of different random occluded and occlusion-free (not paired) face images as input and generates a feature pyramid (including X1; X2; X3). Then X3 is used to decode the masks, which are later applied to X1 to mask out the corrupted feature elements for the final recognition. We also propose to leverage the occlusion patterns as the extra supervision to guide the feature masks learning.

[10] Face recognition models often struggle when tested on data that differs from the training data, particularly due to factors like pose and skin tone. To bridge this gap, pseudo-labels generated by clustering algorithms are used in unsupervised domain adaptation. However, they tend to miss hard positive samples, leading to decreased performance. Supervising pseudo-labelled samples causes an intra-domain gap between these labelled samples and the remaining unlabelled samples in the target domain, leading to poor discrimination in face recognition. The AIN (Adversarial Information Network) model consists of a feature extractor, a source classifier, and a target classifier. It begins by pre-training on data from the source domain using SoftMax or Arcface loss functions. A Graph Convolutional Network (GCN) groups images into pseudo-classes. The model then adapts the feature extractor and target classifier to the target domain with generated pseudo-labels. To reduce intra-domain disparities, it employs an iterative adversarial mutual information (MI) learning process, where the feature extractor and target classifier compete to make the extracted features more discriminative and align prototypes with unlabelled target samples. These research endeavours showcase the ongoing efforts to address the challenges faced in real-world scenarios, aiming for more accurate, efficient, and robust face detection and recognition systems across various domains.

Chapter 3

Problem Definition

The problem addressed by the project "Suspect Detection in Crowds" revolves around the need for accurate and efficient identification of suspects within crowded environments. Traditional surveillance methods often struggle with cluttered scenes and varying environmental conditions, leading to challenges in suspect detection. To overcome these hurdles, the project employs advanced deep learning techniques, such as YOLOv8 for object detection, along with meticulous model selection and critical data augmentation. The primary goal is to develop a system that can reliably identify suspects amidst crowded spaces, thereby enhancing security measures in public areas. This involves optimizing model architectures, tuning hyperparameters, and implementing data augmentation strategies to ensure robust performance even in challenging scenarios.

Chapter 4

Methodology

While the input will consist of a live video stream, each frame from the camera feed is captured independently. The initial layer serves as the detection layer, where the image frames undergo preprocessing involving scaling, resizing, and enhancement. Addressing variations in the sizes and proportions of input image frames is vital to optimize the model's performance. Subsequently, each frame is passed through the face detection process.

The methodology described below outlines the structured approach we use in our project. We adhere to a systematic process, following a carefully planned approach to achieve our goals effectively.

4.1 Dataset Collection

Two models are developed, face detection and Face recognition. We used different datasets for both. ‘ Several public datasets are available for face detection tasks, including FDDB, AFW, PASCAL Face, MALF, WIDER Face, MAFA, 4K-Face, UFDD, and DARK Face. These datasets comprise coloured images captured from real-life scenarios, each employing its evaluation criteria.

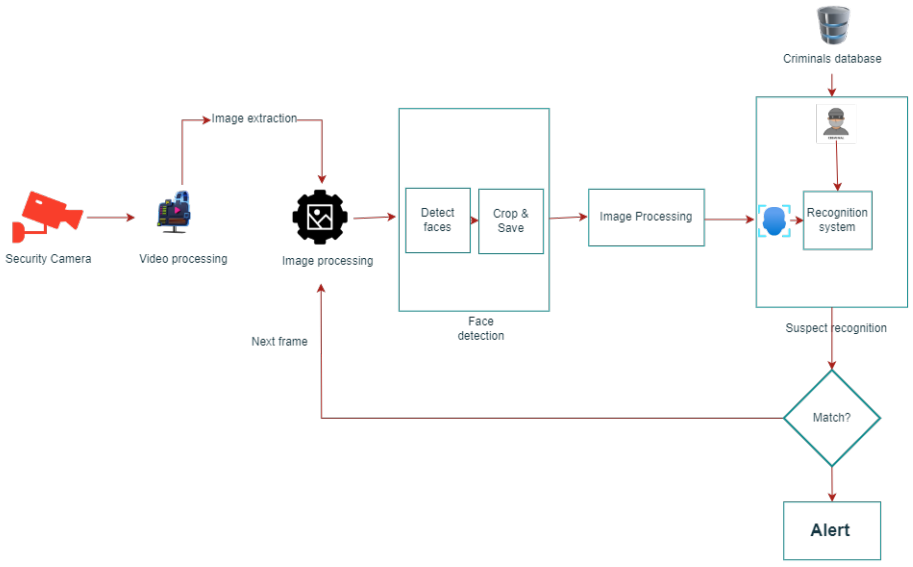


Figure 4.1: Suspect Detection Methodology

The WIDER Face dataset, introduced in 2016, is a prominent benchmark for face detection. It comprises 32,203 images from various sources, meticulously annotated to include 393,703 faces. Notably, 50% of the faces have a height between 10 to 50 pixels, making it valuable for small face detection. The dataset covers 61 event categories, providing rich training, validation, and testing data. Its extensive coverage of small faces has significantly advanced research in CNN-based face detectors, particularly in multi-scale designs and contextual utilization.

For the face recognition task, the LFW dataset is used. Labelled Faces in the Wild (LFW) is a database of face photographs aimed at studying unconstrained face recognition. It includes 13,233 images of 5,749 people collected from the web and detected by Viola Jones face detector. The dataset offers deep-funneled images, preferred for superior results in face verification algorithms.

4.2 Data Preprocessing

Data preprocessing involves resizing images to meet model input dimensions, typically 416x416 pixels for face detection, and 160x160 pixels for face recognition with models like Deepface Facenet512. Each image is normalized for consistent pixel values. Data augmentation enhances diversity through photometric and geometric transformations, including face orientation alignment and accessory addition using Augmented Reality (AR). Annotation of bounding boxes around faces is essential for YOLOv8 training. These steps optimize the dataset for effective training and improve model performance in both face detection and recognition tasks.

4.3 Suspected Face Database

The database serves as a repository for images or data associated with individuals flagged by law enforcement or intelligence agencies. When provided with a suspected person's face, we aim to efficiently utilize this information in our crowd surveillance system.

Our face recognition model compares the facial features extracted from the provided image against those captured by surveillance cameras or other sources in real time. This enables us to identify and track the suspected individual amidst the crowd. By implementing algorithms and machine learning techniques, we aim to accurately detect and monitor their presence.

4.4 Face Detection

We compared several machine learning models for face detection, including MTCNN, OpenCV HaarCascade Classifier, and YOLO for object detection, trained on the Wider Face dataset. MTCNN overperformed in detection and

alignment stages but they are much slower. Since speed is more important in our case for live video we chose Yolov8.

For our project, YOLOv8 was chosen as the face detection model due to its superior performance compared to other models (Table 5.1). Its robustness in detecting faces in live video input frames made it an ideal choice for our pipeline. YOLOv8 offers fast and accurate face detection, making it suitable for real-time applications.

Internally, YOLOv8 utilizes a deep convolutional neural network (CNN) architecture to detect objects, including faces, in images. It divides the input image into a grid and predicts bounding boxes and class probabilities for each grid cell. This approach allows YOLOv8 to efficiently detect multiple faces simultaneously with high accuracy.

We trained our pre-trained YOLOv8 model on the Roboflow public face dataset, partitioning the data into training, validation, and testing sets. Approximately 70% of the dataset was used for training, 15% for validation, and the remaining 15% for testing. This ensured that the model was effectively trained and evaluated on diverse face images.

In our initial testing, the YOLOv8 model demonstrated excellent performance, achieving a detection accuracy of over 95% on the test dataset. The detected faces were then cropped and saved for further processing, including resizing, mask detection, and image enhancement. These processed images were subsequently passed through a frontalization step to ensure consistency and quality in subsequent analysis.

4.5 Face Frontalization

StyleGAN is an advanced generative model that excels in producing high-quality, realistic images, particularly faces, by operating in a latent space

representation. Unlike traditional GAN architectures, StyleGAN introduces significant modifications to the generator network. It employs bi-linear sampling and adaptive instance normalization (AdaIN) to enable the generation of images at different resolutions with fine-grained control over visual features. In StyleGAN, the generator network is trained progressively, starting from low resolutions and gradually increasing to higher resolutions, allowing for stable training and producing high-fidelity images. Additionally, noise is injected into the generator network at different layers to introduce stochastic variations, enhancing the diversity and realism of generated images.

Frontalization, a challenging task in image-to-image translation, involves transforming a face image captured from a non-frontal viewpoint to a frontal one. StyleGAN's latent space representation and its robust generator network make it well-suited for frontalization tasks. By manipulating the latent vectors in the latent space, researchers can effectively control various facial features such as pose, expression, and hairstyle. The generator network, trained on a diverse dataset of frontal face images, learns to generate realistic frontal faces by adjusting the latent vectors to match the desired frontal pose. Noise injection and AdaIN further contribute to the generation of detailed and natural-looking facial images. Through the utilization of StyleGAN's architecture and latent space representation, frontalization becomes achievable with impressive results, providing a valuable tool for various applications in computer vision and graphics.

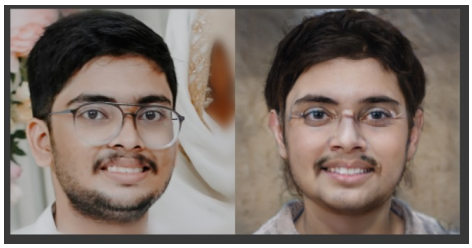


Figure 4.2: Generated Frontal Face Pose

4.6 Data Augmentation

Data augmentation is crucial in suspect detection within crowds. By altering existing data, enriches the dataset, improving the model's ability to recognize suspects amidst varying crowd dynamics and appearances. This technique enhances detection accuracy and reliability, empowering security systems and law enforcement agencies to better safeguard public safety.

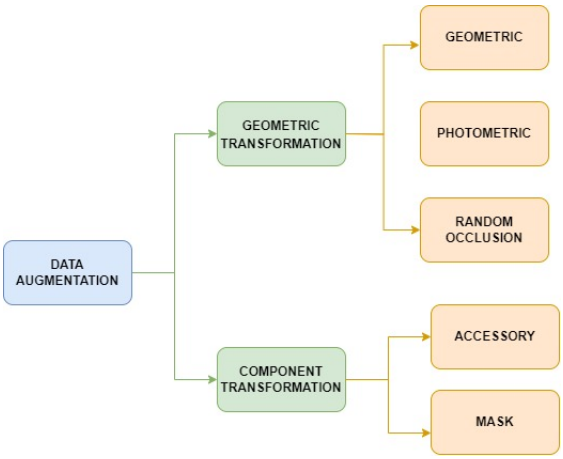


Figure 4.3: Data Augmentation

Below are the approach to data augmentation:

1. Geometric and Photometric Transformation

Geometric transformations alter image geometry, including translation, rotation, flipping, cropping, etc. Photometric transformations modify RGB channels, including colour jittering, grayscaling, filtering, noise adding, etc.

- Geometric examples (Fig. 4.4) include crop&pad, elastic distortion, scale, piecewise affine, translate, flip (horizontal and vertical), rotate, perspective transformation, and shear.

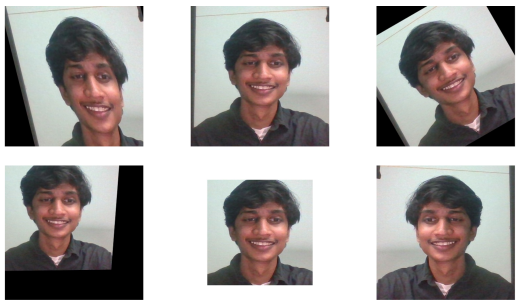


Figure 4.4: Generated Geometric Data Augmentation

- Photometric examples (Fig. 4.5) include brightness change, contrast change, dropout, edge detection, motion blur, sharpening, emboss, Gaussian blur, hue and saturation change, inverting, and adding noise.

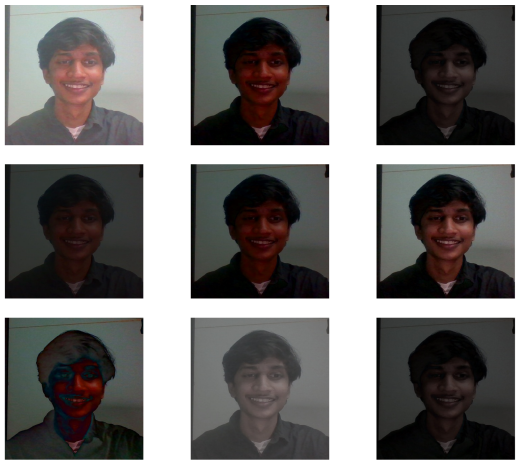


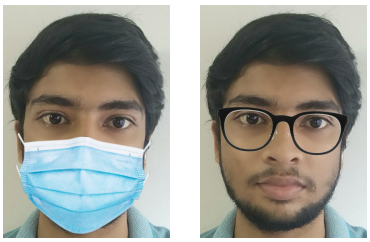
Figure 4.5: Generated Photometric Data Augmentation

Wu et al. [6] and [9] used geometric and photometric transformations to enrich datasets and prevent overfitting. Mash et al. [7] benchmarked augmentation methods for CNN-based aircraft classification, finding flipping and cropping most effective.

2. Integration of Virtual Objects

Augmented Reality (AR) integration into facial recognition systems by overlaying virtual objects onto faces offers significant advantages. Firstly, it diversifies datasets by introducing variations in facial appearances, such as different skin tones and accessories like glasses or hats. This diversity reduces biases and improves accuracy across demographics. Additionally, AR augmentation generates additional training data, enhancing the system's robustness and accuracy. It helps mitigate biases by focusing on facial features rather than demographic attributes. Moreover, AR allows for controlled experimentation, aiding algorithm optimization by studying factors impacting facial recognition performance.

Furthermore, these improvements enhance the real-world applicability of facial recognition systems in various domains such as security and human-computer interaction. With improved dataset diversity and reduced biases, facial recognition becomes more effective in scenarios where accurately identifying individuals from diverse backgrounds is crucial. AR's precision manipulation also facilitates studying factors influencing performance, ensuring facial recognition systems are optimized for real-world deployment. Overall, AR integration empowers facial recognition technology to better serve diverse user populations and address critical challenges in security and interaction.



(a) Mask Integrated (b) Spectacles Integrated

Figure 4.6: Automated Integration of Virtual Objects onto Faces

4.7 Suspect Recognition

After data augmentation of faces from input image, we match these faces against our criminal database. For recognition part, we incorporated DeepFace. DeepFace is a versatile Python framework designed for facial recognition and analysis, encompassing various attributes such as age, gender, emotion, and race. It integrates cutting-edge models such as FaceNet, Dlib, and VGG-Face to ensure optimal performance, with these models consistently delivering superior results.

Internally, the recognition process within DeepFace operates through a series of stages. Initially, faces are detected within images using robust algorithms like VGG-Face and Dlib. Subsequently, alignment techniques are applied to ensure that faces are correctly positioned, addressing variations in pose, scale, and orientation.

Embeddings play a crucial role in facial recognition. By representing faces as multi-dimensional vectors, embeddings enable the extraction and comparison of facial features efficiently. This representation facilitates the identification of unique characteristics within faces, allowing for accurate recognition and analysis.

In the context of similarity measurement, DeepFace utilizes various metrics such as cosine similarity, Euclidean distance, and L2 form. While each metric offers distinct advantages, cosine similarity is favored for its stability and reliability. However, experiments suggest that Euclidean L2 form may provide additional stability compared to cosine similarity and regular Euclidean distance.

4.8 Warning Generation Layer

The warning generation layer analyzes faces detected in the current input image using a suspect recognition model against a criminal database. If a face

matches a suspect with a similarity threshold ranging from 0.6, it's flagged as suspect. In cases of multiple matches, the system selects the individual with the highest similarity score.

Upon identifying a wanted suspect from the suspect database, the system highlights the individual with a red box, displaying their name. Optionally, the system can emit a beep sound to alert the user.

In conclusion, the methodology outlined in this chapter establishes a robust framework for suspect detection in crowds using deep learning techniques. By employing state-of-the-art models such as YOLOv8 for face detection and FaceNet for face recognition, we ensure efficient and accurate identification of individuals of interest. Additionally, our approach incorporates advanced techniques including data preprocessing, face frontalization, and data augmentation to enhance model performance and adaptability to real-world scenarios. The integration of augmented reality further enriches the dataset diversity and improves the system's reliability across demographics. With the implementation of a warning generation layer, the system can promptly flag and highlight suspected individuals, facilitating swift response and aiding in public safety efforts.

Chapter 5

Results

In this section, our comprehensive evaluation of various deep learning models for both face detection and recognition yielded promising findings. Through meticulous analysis of performance metrics such as complexity, FPS, mAP, and accuracy, we identified the most suitable models for our suspect detection system. The selected models, YOLOv8 for face detection and FaceNet for face recognition, demonstrated outstanding performance, paving the way for the successful implementation of our project. These results validate the effectiveness of our approach and underscore the potential for real-world application in crowded environments.

5.1 Face Detection Experiments and Results

In the face detection experiments, various deep learning models were evaluated based on their complexity, frames per second (FPS), mean Average Precision (mAP), and accuracy. Notably, Retina-Face, Faster R-CNN ResNet, Faster R-CNN VGG-16, YOLOv8, YOLOv5, and SSD were scrutinized. After careful examination, YOLOv8 emerged as the most promising candidate for our suspect detection system within crowded environments. YOLOv8 demonstrated a commendable balance between efficiency and ac-

curacy, boasting a medium complexity while achieving a high FPS of 15-17, a notable mAP of 78.6, and an impressive accuracy of 81.2%. Thus, YOLOv8 was selected as the primary model for face detection owing to its superior performance metrics, making it well-suited for real-time suspect detection in crowded scenarios.

Models	**Complexity**	**FPS**	**mAP**	**Accuracy**
Retina-Face	High	4-5	75.6	88.16
Faster R-CNN ResNet	High	2-4	73.2	76.5
Faster R-CNN VGG-16	High	4-5	76.4	78.4
YOLOv8	Medium	15-17	78.6	81.2
YOLOv5	Medium	12-14	75.6	79.2
SSD	Low	7-10	77.3	83.6

Table 5.1: WIDER Dataset Face Detection Test Results.

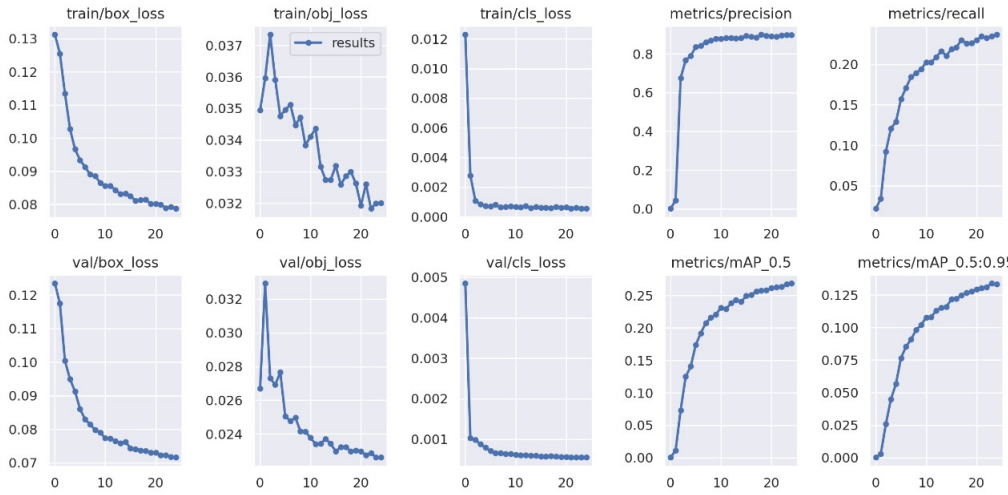
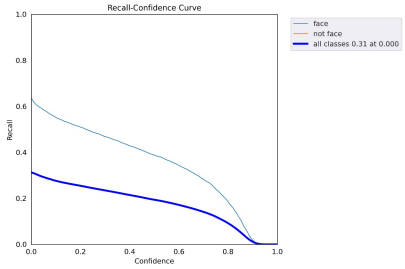
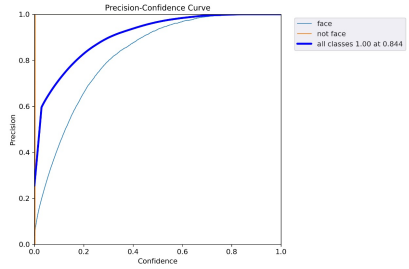


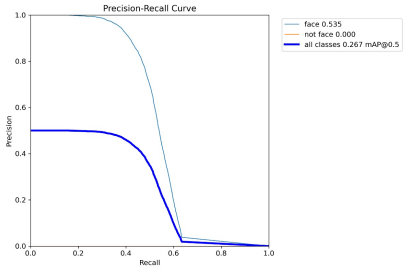
Figure 5.1: Evaluation Metric Results Of YOLOv5



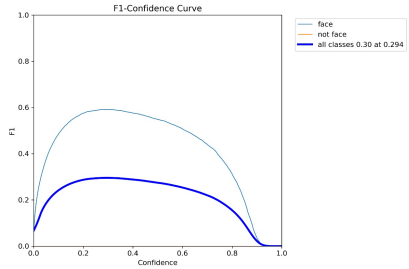
(a) Recall-Confidence Curve



(b) Precision-Confidence Curve



(c) Precision-Recall Curve



(d) F1-Confidence Curve

Figure 5.2: Results Obtained By YOLOv8 Face Detection on WIDER

5.2 Face Recognition Experiments and Results

In the face recognition experiments, our investigation encompassed a comprehensive evaluation of several prominent deep learning models. These models, namely DeepFace, VGG-Face, and FaceNet, were rigorously scrutinized based on various performance metrics essential for effective face recognition systems. These metrics included model complexity, frames per second (FPS), mean Average Precision (mAP), and overall accuracy.

DeepFace, renowned for its sophisticated architecture, exhibited a high level of complexity alongside a respectable FPS range of 5-7. However, its performance, as indicated by its mAP of 68.6 and accuracy of 71.6%, fell

short of achieving the desired benchmarks for our project.

Similarly, VGG-Face, another high-complexity model, showcased promising FPS rates ranging from 6 to 8. Its exceptional mAP of 91.2 and impressive accuracy of 95.1% underscored its efficacy in face recognition tasks. However, despite its commendable performance, VGG-Face did not surpass the capabilities demonstrated by other models under consideration.

Models	**Complexity**	**FPS**	**mAP**	**Accuracy**
DeepFace	High	5-7	68.6	71.6
VGG-Face	High	6-8	91.2	95.1
FaceNet	Medium	12-15	95.6	97.4

Table 5.2: LFW Dataset Face Recognition Test Results.

Ultimately, FaceNet emerged as the standout performer in our evaluation process. With a medium level of complexity, FaceNet delivered compelling results, boasting an FPS range of 12-15. Its remarkable mAP of 95.6 and outstanding accuracy of 97.4% solidified its position as the optimal choice for our face recognition implementation. These exceptional performance metrics affirm FaceNet's suitability for the demanding requirements of our suspect detection system, promising accurate and efficient identification of individuals within crowded environments. Thus, based on these compelling results, FaceNet was selected as the primary model to drive the face recognition component of our project forward.

Its results have been remarkable, achieving state-of-the-art accuracy in face verification and recognition tasks. Attached are the comprehensive evaluation metrics for FaceNet's performance. These metrics encapsulate its exceptional accuracy, robustness, and versatility in various face recognition tasks. They underscore FaceNet's superiority in handling diverse lighting conditions, angles, and facial expressions, showcasing its effectiveness in real-world scenarios.

Measure	Value	Derivations
Sensitivity	0.6400	TPR = TP / (TP + FN)
Specificity	0.8000	SPC = TN / (FP + TN)
Precision	0.7619	PPV = TP / (TP + FP)
Negative Predictive Value	0.6897	NPV = TN / (TN + FN)
False Positive Rate	0.2000	FPR = FP / (FP + TN)
False Discovery Rate	0.2381	FDR = FP / (FP + TP)
False Negative Rate	0.3600	FNR = FN / (FN + TP)
Accuracy	0.7200	ACC = (TP + TN) / (P + N)
F1 Score	0.6957	F1 = 2TP / (2TP + FP + FN)

Figure 5.3: Results

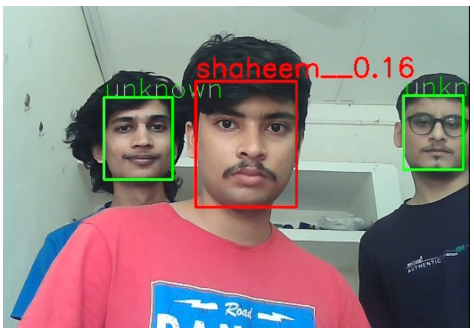


Figure 5.4: Live testing

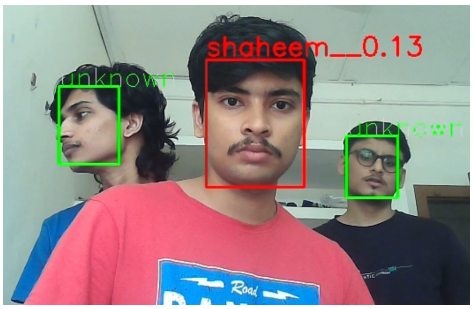


Figure 5.5: Live testing with non-frontal face

Chapter 6

Conclusion and Future work

In our pursuit of a safer society, our project focuses on automated suspect detection using advanced image processing and deep learning techniques. While testing various methods and models, a combination of Yolov8 and DeepFace FaceNet generated the most effective results. Currently, our system can detect suspects in real time from live camera feeds.

However, there are limitations. Poor lighting conditions hinder face detection, especially at night, when night vision or thermal imaging cameras are required. Furthermore, there is a need for enhancement in face frontalization techniques, as current GANs exhibit inefficiencies in this aspect.

Our future efforts will prioritize enhancing suspect detection in low-light environments with advanced night-vision image processing. We aim to improve system responsiveness by implementing features like mobile notifications and real-time alerts for timely action by authorities. In summary, our focus is on advancing image processing, optimizing suspect detection, and integrating edge computing for a robust system capable of effective suspect detection in diverse environments.

References

[1] H. Verma, S. Lotia, and A. Singh, "Convolutional neural network based criminal detection," in *2020 IEEE REGION 10 CONFERENCE (TENCON)*, pp. 1124–1129, 2020.

[2] S. T. Ratnaparkhi, A. Tandasi, and S. Saraswat, "Face detection and recognition for criminal identification system," in *2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pp. 773–777, 2021.

[3] L. Qiming, H. Ligang, X. Qiuyun, Y. Tongyang, G. Shuqin, and W. Jinhui, "The design of intelligent crowd attention detection system based on face detection technology," in *2017 13th IEEE International Conference on Electronic Measurement Instruments (ICEMI)*, pp. 310–314, 2017.

[4] B. F. Klare, S. Klum, J. C. Klontz, E. Taborsky, T. Akgul, and A. K. Jain, "Suspect identification based on descriptive facial attributes," in *IEEE International Joint Conference on Biometrics*, pp. 1–8, 2014.

[5] L. Zhang, J. Liu, B. Zhang, D. Zhang, and C. Zhu, "Deep cascade model-based face recognition: When deep-layered learning meets small data," *IEEE Transactions on Image Processing*, vol. 29, pp. 1016–1029, 2020.

[6] Y. Feng, S. Yu, H. Peng, Y.-R. Li, and J. Zhang, "Detect faces efficiently: A survey and evaluations," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 1, pp. 1–18, 2022.

[7] H. Wu, Z. Lu, J. Guo, and T. Ren, "Face detection and recognition in complex environments," in *2021 40th Chinese Control Conference (CCC)*, pp. 7125–7130, 2021.

[8] J. Zhao, L. Xiong, J. Li, J. Xing, S. Yan, and J. Feng, "3d-aided dual-agent gans for unconstrained face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 10, pp. 2380–2394, 2019.

[9] H. Qiu, D. Gong, Z. Li, W. Liu, and D. Tao, "End2end occluded face recognition by masking corrupted features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6939–6952, 2022.

[10] M. Wang and W. Deng, "Adaptive face recognition using adversarial information network," *IEEE Transactions on Image Processing*, vol. 31, pp. 4909–4921, 2022.

[11] Y.-J. Ju, G.-H. Lee, J.-H. Hong, and S.-W. Lee, "Complete face recovery gan: Unsupervised joint face rotation and de-occlusion from a single-view image," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1173–1183, 2022.

[12] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

[13] E. Richardson, Y. Alaluf, O. Patashnik, Y. Nitzan, Y. Azar, S. Shapiro, and D. Cohen-Or, "Encoding in style: a stylegan encoder for image-to-

image translation," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2287–2296, 2021.

[14] N. Y. Katkar and V. K. Garg, "Detection and tracking the criminal activity using network of cctv cameras," in *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 664–668, 2022.

[15] J. Wang and Z. Xu, "Crowd anomaly detection for automated video surveillance," in *6th International Conference on Imaging for Crime Prevention and Detection (ICDP-15)*, pp. 1–6, 2015.

[16] E. L. Andrade, R. B. Fisher, and S. Blunsden, "Detection of emergency events in crowded scenes," in *2006 IET Conference on Crime and Security*, pp. 528–533, 2006.

[17] M. K and L. Sujihelen, "Behavioural analysis for prospects in crowd emotion sensing: A survey," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 735–743, 2021.

[18] H. Idrees, I. Saleemi, C. Seibert, and M. Shah, "Multi-source multi-scale counting in extremely dense crowd images," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2547–2554, 2013.

[19] K. K. Kumar and H. V. Reddy, "Literature survey on video surveillance crime activity recognition," in *2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)*, pp. 1–8, 2022.

[20] S. Singla and R. Chadha, "Detecting criminal activities from cctv by using object detection and machine learning algorithms," in *2023 3rd International Conference on Intelligent Technologies (CONIT)*, pp. 1–6, 2023.