Suspect Detection from Crowd Using Deep Learning

Mohammed Shaheem C, Shreyas S N, Vinit Kumawat
Under the guidance of
Dr Santosh Kumar Behera
Assistant Professor – CSED
National Institute of Technology Calicut

***Abstract*- Along with the fast-growing economy of a developing country as far as India is concerned, an emerging number of crimes and criminal offenses 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.**

Index

I. Introduction

In today's security-conscious world, deep learning [1] offers an opportunity to enhance suspect detection in crowded places. This research explores the integration of deep learning techniques to improve suspect identification. We focus on facial recognition and witness-based suspect identification using deep learning methodologies.

Deep learning methods, like MTCNN for detection and FaceNet for embeddings [2], play a crucial role in achieving accurate and non-intrusive criminal identification [2]. We also delve into crowd attention analysis, using deep learning to create an intelligent system for real-time monitoring [3].

Our research addresses the need for an efficient suspect detection system that can perform well in diverse conditions, including dynamic datasets and imperfect images. We aim to bridge gaps left by previous research, ensuring our system functions even under challenging conditions. Detailed examinations of referenced papers will be presented in the Literature Survey section.

II. Motivation

The need for a strong security system, capable of recognizing individuals in challenging situations, is vital for crime prevention, reducing repeat offenses, improving public safety, aiding investigations, and maintaining national security. This plays a crucial role in upholding law and order and ensuring the well-being of society.

On the other hand, existing security systems face various issues. These problems include reduced low-light performance, difficulties with people turning their heads, and challenges with obstructions like scarves, face masks, and sunglasses. Some systems also struggle when dealing with data they haven't seen before. Moreover, criminals actively try to hide by avoiding direct camera views and adding obstructions near their faces, making it harder for the system to recognize them.

Regarding high-resolution cameras, although they offer clear images and fine details, they come with their own set of problems. They are expensive, demand a lot of data and maintenance, and are vulnerable to outdoor conditions. Therefore, achieving a balance in outdoor security, a combination of advanced technology and practical considerations, is crucial to effectively address these complex challenges.

III. 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 emphasising male images. The VGG CNN models achieve a remarkable 99.5% 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 utilises 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 on 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. Labelling 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 artifacts.

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 auto-encoder as the discriminator. The dual agents are responsible for discriminating real versus fake (minimizing the adversial 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 (Adversial 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) group 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.

IV. Methodology

In this research, we propose a methodology for suspect detection in crowded environments [3] [7] using deep learning [1]. The process begins with acquiring input images, which have been pre-processed and subjected to feature extraction. These features are then stored in a database for subsequent matching with the facial images captured by video surveillance systems. Detecting faces within the video footage is a crucial step, and subsequent preprocessing is conducted to extract discriminative features that aid in identifying potential suspects in complex and crowded environments.t

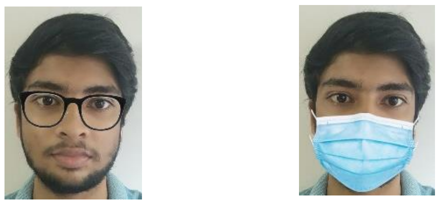
Goal:

We aim to develop a robust deep learning-based framework that overcomes challenges such as inconsistent face proportions due to varying camera distances, occlusion, mutual face blocking within crowds, and low-resolution facial images. The aim is to design a system that effectively detects suspects in complex and crowded environments, enhancing overall security and public safety.

Comprehensive Data Collection: Our dataset encompasses diverse samples, notably the Roboflow public face detection dataset, which comprises over 10,000 faces. This extensive dataset enables us to train our model to detect faces across varied conditions, including different angles, poses, occlusions, and varying resolutions. Furthermore, we supplement this dataset with additional sources to ensure comprehensive coverage of potential scenarios encountered in crowded environments.

Face Frontalization with GANs: To ensure optimal recognition accuracy, we utilize Generative Adversarial Networks (GANs) to frontalize detected faces. Frontalization enhances recognition probabilities, especially as frontal poses tend to facilitate more accurate recognition. This step is critical in mitigating the challenges posed by varying poses and angles, ensuring that our model can effectively identify suspects regardless of their orientation within the crowded scene.

Data Augmentation Techniques: Employing a range of augmentation techniques, we enhance the dataset further. These techniques include geometric and photometric transformations, random occlusion, and component and attribute alterations such as hairstyle, makeup, pose, expression, and age. Augmentation enriches the dataset, thereby improving recognition accuracy and resilience against variations encountered in crowded environments. Moreover, we continuously refine and expand our augmentation strategies to encompass a wider range of potential variations, ensuring the robustness of our model in real-world scenarios.



Samples of data augmented

Training with Crystal Loss Function: Our model is trained using the Crystal Loss function, specifically designed for face verification and recognition systems. This loss function constrains features to lie on a hypersphere of a fixed radius, optimizing similarity scores for positive pairs while minimizing scores for negative pairs. This refinement significantly enhances the model's performance, ensuring robust suspect detection capabilities. Additionally, we employ techniques such as transfer learning and fine-tuning to adapt pre-trained models to our specific task, leveraging existing knowledge to improve efficiency and effectiveness.

Real-Time Suspect Detection and Alert System: Upon detection of a suspected individual by the surveillance camera, an immediate alert is dispatched to nearby security personnel, facilitating swift intervention. This real-time alert system aims to enhance safety and security in crowded environments effectively. Furthermore, we continuously evaluate and refine the alert system to minimize false positives and ensure timely and accurate notifications, optimizing the response to potential threats in crowded settings.

Future Directions: Moving forward, our research will extend beyond suspect detection to encompass monitoring crowd behavior. This includes detecting incidents such as fights, potential terrorist activities, pickpocketing, and other relevant scenarios. By broadening the scope of our model's capabilities, we aim to further enhance safety measures in crowded settings. In conclusion, our methodology integrates advanced deep learning techniques with robust datasets to elevate suspect detection capabilities in crowded environments. We anticipate that our research will make significant contributions to the field of surveillance and security, fostering safer environments for all. Implementation of a real-time alert system facilitates swift response to potential threats, with continual refinement ensuring adaptability and optimization for enhanced security measures.

V. Conclusion

Following an extensive literature review and comprehensive analysis of suspect detection in crowded areas, we have identified several prominent methodologies. However, various recommendation models proposed by previous researchers encounter notable challenges. Our main objective is to formulate effective solutions to address these issues.

In pursuit of our forward-thinking goals, we have outlined a structured work plan. This plan encompasses continuous literature surveys to stay up-to-date with the latest advancements, coupled with a dedicated emphasis on gaining expertise in advanced Python libraries and deep learning concepts. Mastery of these elements is paramount for the success of our project and will underpin the development of our preliminary solution framework.

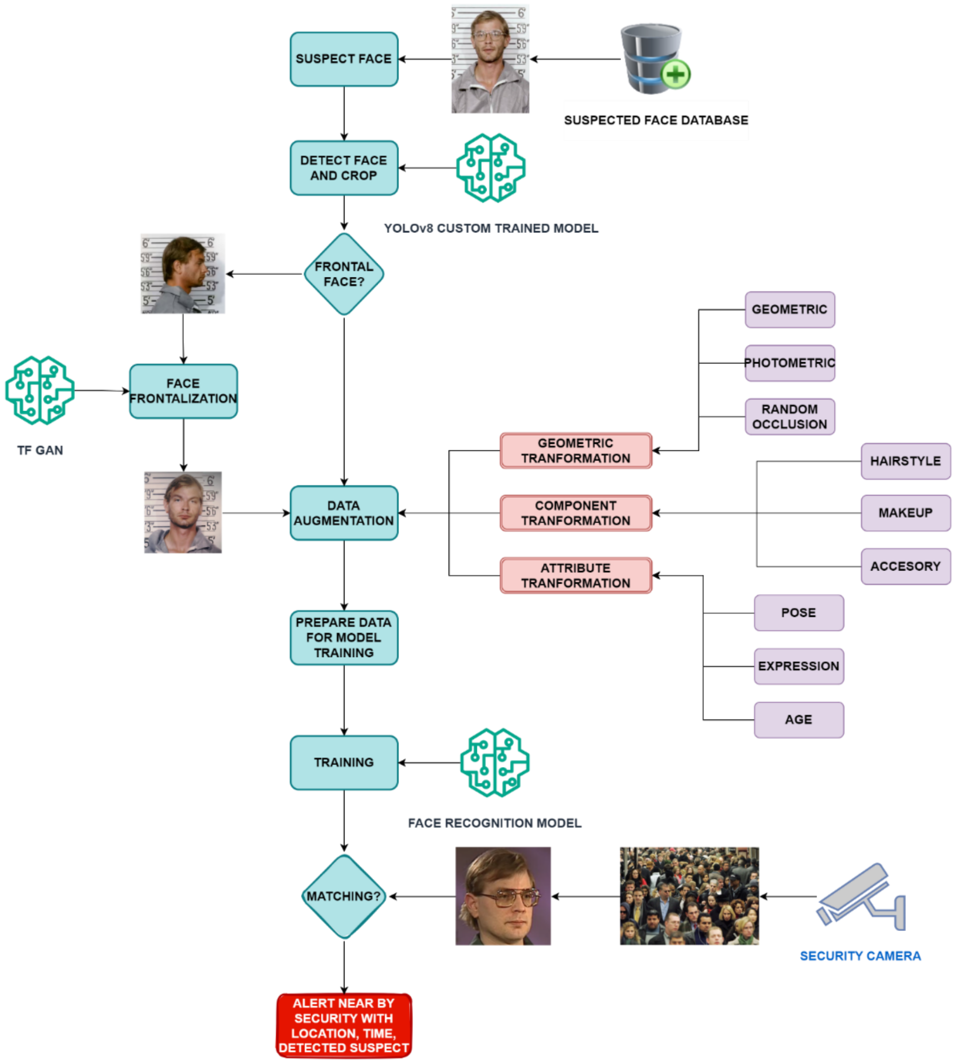


Figure 1: suspect detection methodology.

References

[1] H. Verma, S. Lotia and A. Singh, "Convolutional Neural Network Based Criminal Detection," IEEE REGION 10 CONFERENCE (TENCON), Osaka, Japan, 2020.

[2] S. T. Ratnaparkhi, A. Tandasi and S. Saraswat, "Face Detection and Recognition for Criminal Identification System," 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 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," 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), Yangzhou, China, 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," IEEE International Joint Conference on Biometrics, Clearwater, FL, USA, 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, 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, Jan. 2022.

[7] H. Wu, Z. Lu, J. Guo and T. Ren, "Face Detection And Recognition In Complex Environments," 40th Chinese Control Conference (CCC), Shanghai, China, 2021, pp. 7125-7130, doi: 10.23919/CCC52363.2021.9549933.

[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, 1 Oct. 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, 1 Oct. 2022, doi: 10.1109/TPAMI.2021.3098962.

[10] M. Wang and W. Deng, "Adaptive Face Recognition Using Adversarial Information Network," IEEE Transactions on Image Processing, vol. 31, pp. 4909-4921, 2022.