www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 12 Issue 12 December 2023, Page No. 25965-25979
ISSN: 2319-7242 DOI: 10.18535/ijecs/v12i11.4781

Incorporating MTCNN and DeepFace: A Novel Hybrid Deep Learning Method for Human Face Detection and Recognition

Honey, Dr. Sukhwinder Singh Oberoi

Punjabi University Patiala

Abstract:
The realm of face detection has become a focal point of extensive research, driven by its diverse applications spanning computer vision, communication, and automatic control systems. Realizing real-time recognition of multiple faces within embedded systems poses a formidable challenge due to the intricate computational demands involved. This challenge necessitates a deep exploration of facets such as face detection, expression recognition, face tracking, and pose estimation. Accurately identifying a face from a single image stands as the core challenge, primarily due to the non-rigid nature of faces, resulting in variations in size, shape, color, and more. Furthermore, the complexity of face detection amplifies when confronted with unclear images, occlusions, suboptimal lighting conditions, off-angle poses, and various other factors. This study presents an innovative framework for multiple face recognition. Through extensive experiments, the system's prowess in simultaneously recognizing up to 10 different human face poses in real time was showcased, achieving remarkable processing speeds as low as 0.21 seconds. The system demonstrated an impressive minimum recognition rate of 93.15%, underscoring the effectiveness of the proposed methodology. While the primary emphasis lies on frontal human faces, the system is adept at handling poses beyond the frontal orientation, marking a significant advancement in the domain of face detection and recognition.

Keywords: Face Recognition, Face Detection, DeepFace, FaceNet, Computer Visions.

I. Introduction

With the advent of deep learning, the field of computer vision has undergone a significant transformation, attracting considerable interest. Among the challenging problems addressed, face recognition stands out as a topic that has garnered extensive exploration. The utilization of face recognition extends to numerous applications in both private and public domains, including surveillance systems and person authentication. Traditional face recognition methods have been geared towards the development of quicker and more resilient algorithms [1]. However, achieving accurate recognition hinges on obtaining high-resolution, unobstructed facial images that possess discriminative qualities for identity changes while remaining resilient to intra-personal variations. Despite the advancements in face recognition algorithms, their accuracy is still influenced by external and internal factors such as illumination, pose, facial expressions, and occlusions. A primary hurdle in face recognition lies in creating effective feature representations that enhance accuracy across diverse scenarios [2]. Recent strides in deep learning have yielded substantial progress in computer vision research, particularly in classification and recognition tasks. Researchers have introduced deep learning-based face recognition algorithms that showcase commendable performance in terms of both processing speed and accuracy. Yet, the real-time processing of multiple face recognition remains a challenge for deep learning algorithms due to their high computational complexity. In typical face recognition algorithms, a raw facial image is fed into a neural network for learning facial features through Convolutional Neural Networks (CNN), pooling, and fully connected layers. Each step contributes to substantial computational resource demands. This paper introduces a novel framework for real-time multiple face recognition. The model employs a network architecture that reduces network parameters, and incorporates tracking techniques to accelerate processing times while maintaining an acceptable recognition rate [3]. Image processing involves employing tools or algorithms to manipulate images with the aim of tasks such as image compression, image enhancement, or extracting valuable information from the image. This process can be approached through two methods: digital and analog. Digital image processing employs mathematical models and computer algorithms to process digital images, while analog image processing pertains to handling hard copies like photographs and printed material. Within the realm of image processing lies the field of face detection. In recent times, the study of face detection has gained significance due to the pervasive integration of computers in various aspects of human interaction. One of the domains where this holds importance is video surveillance, which is employed for security purposes [4]. Video surveillance encompasses multiple phases and plays a crucial role in ensuring safety.

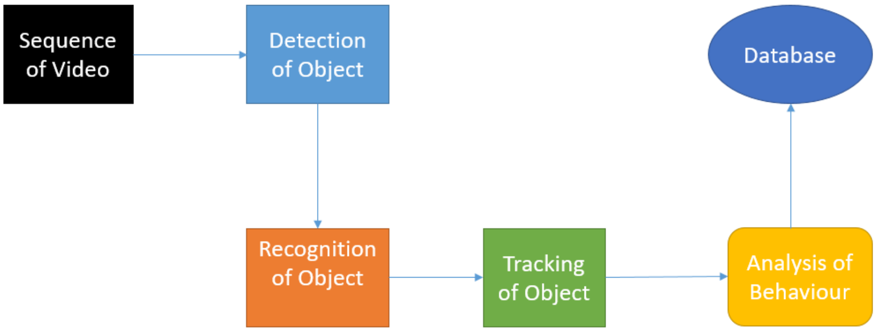

Figure 1: Phases of Video Surveillance

In the realm of face detection, two types of errors are recognized: false negatives and false positives. A false positive occurs when an image region is classified as lacking a face even though a face is present. Conversely, a false negative arises when faces are overlooked due to a low detection rate. In algorithmic evaluation, various measures are employed, including detection rate, learning time, execution time, and the ratio between detection rate and false values [5]. The detection rate denotes the proportion of human-detected faces relative to the number of faces correctly identified by the system. Within this research field, specific terms pertaining to faces hold significance. “Facial feature detection” refers to pinpointing distinct human facial features such as eyes, nose, lips, chin, eyebrows, mouth, and ears. “Face authentication” involves verifying a person’s identity based on their face. “Face tracking” entails real-time monitoring of a face’s position. “Expression recognition” encompasses the identification of facial expressions such as happiness, sadness, surprise, and more. Lastly, “face localization” concerns the determination of a face’s precise location within an image.

The face detection process involves three key components. Firstly, the introduction of an integral image, also known as the integral image, enables rapid computations for the detector. Secondly, a learning algorithm based on the MTCNN technique is employed. MTCNN is a sophisticated facial detection algorithm, renowned for its capability to accurately identify faces within images. This cutting-edge technology employs a multi-stage process to locate facial features, making it a robust choice for various applications such as facial recognition, emotion analysis, and even security systems [6]. Its ability to efficiently detect faces in images has made it a crucial tool in computer vision and AI-driven solutions. This algorithm swiftly identifies the most prominent features from an extensive feature set, resulting in an exceptionally efficient classifier. Lastly, a method is used to combine more intricate classifiers into a cascade, expediting the elimination of image backgrounds. Effective face detection hinges on the clarity of the integral image. Specifically, the resolution of the input image must exceed 5x5 resolutions in order to achieve successful face detection.

Related Work

While face recognition comes naturally to humans, it presents a more challenging task for computers due to the non-rigidity of the human body. This non-rigidity leads to continuous changes over time, making consistent recognition difficult. The foremost challenge in face recognition is the initial step of face detection. This serves as a crucial prerequisite for automated face recognition systems. In the process of face detection, various complexities arise, such as differing facial orientations yielding distinct appearances. To address this issue, a strategy involves curating a collection of diverse face images, each representing a distinct positional viewpoint for a given individual. Moreover, variations in facial expression and lighting conditions contribute to disparities in face images even when the expression and position are constant [7]. This paper primarily revolves around a frontal face detection system and Tilted faces, which is capable of identifying faces. However, it's noteworthy that this system may have a certain rate of false positives, where non-faces are erroneously identified as faces. In face recognition, the capability to accurately identify tilted or non-frontal faces is a crucial aspect. While the primary emphasis often lies on recognizing faces in a frontal pose, modern face recognition systems are designed to handle faces with varying orientations. These systems employ advanced algorithms, such as pose estimation and landmark detection, to infer the pose of a face and align it properly. This ensures that the recognition process remains effective even when faces are turned or tilted. The ability to handle tilted faces enhances the applicability of face recognition across real-world scenarios, including surveillance, crowd analysis, and human-computer interaction.

Within this section, we will initially outline the conventional (Deep Learning) algorithm employed within the framework. The pivotal components of this framework can be segmented into three main sections, namely face detection and alignment, face recognition, and tracking algorithm. These components are illustrated in figure 1. The ensuing discussion will delve into the literature encompassing each of these distinct facets.

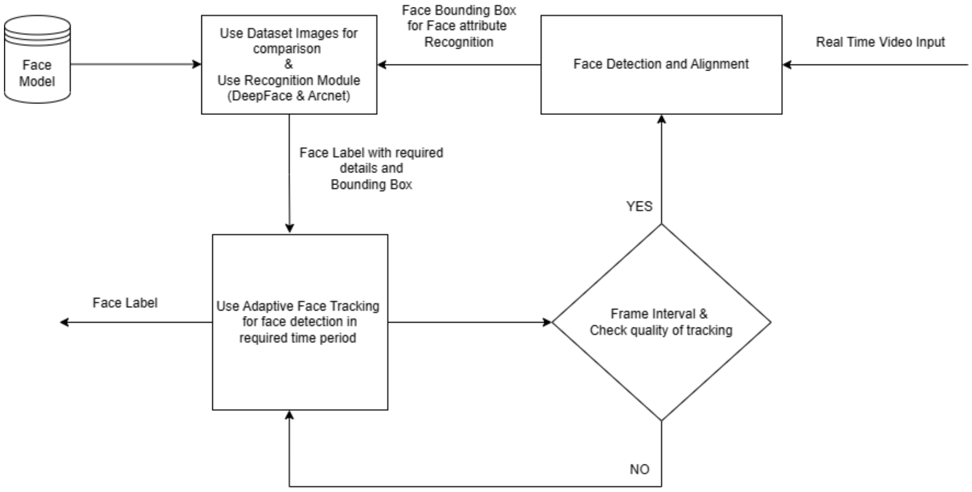

Figure 2: Blueprint of Proposed Framework

Knowledge-Based Approach: This method operates on a set of rules, seeking to establish correlations among various features of the human face through pre-existing knowledge. It relies on encoded understandings of facial attributes to delineate the constituents of a human face, making it applicable for localizing human faces.

Feature-Invariant Approach: Employed for face localization, these techniques are designed to identify faces across diverse conditions, encompassing scenarios with varying levels of light intensity, facial poses from multiple angles, and distinct viewpoints. By amalgamating these diverse conditions, this approach achieves the ability to detect faces under various circumstances, offering a solution to the challenge posed by different facial poses.

Template-Based Approach: The template matching technique serves a dual purpose, functioning as both a means of localizing and detecting faces. This method involves storing various predefined patterns representative of human faces, which can encapsulate either the entire face or specific facial features. The objective is to establish a correlation between the stored templates and the input image, facilitating the detection process.

Appearance-Based Approach: Within this methodology, the emphasis shifts from templates to a compilation of training images. These training images are employed to construct a model of the image, which is subsequently harnessed for the purpose of face detection.

1. Face Detection and Alignment:

Face detection constitutes a pivotal phase within the framework, serving as a fundamental input to the subsequent recognition process. In this context, we implemented a face detection algorithm founded upon, leveraging a Deep Learning (MTCNN) approach [8]. This algorithm showcases the capacity to detect faces across diverse variations and lighting conditions. The resulting output of the detection process encompasses both a bounding box delineating the face's location and facial landmarks that facilitate precise alignment. The algorithm initiates by resizing the input image into a range of scales, effectively creating an image pyramid. MTCNN, short for Multi-Task Cascaded Convolutional Networks, is a state-of-the-art deep learning framework used for face detection tasks. It employs a cascade of three neural networks to detect faces with varying levels of complexity. The first network, the Proposal Network (P-Net), generates a set of candidate bounding boxes around potential faces. These boxes are then refined and filtered by the following two networks. The second network, the Refine Network (R-Net), further refines the candidate boxes and eliminates false positives [9]. It also performs facial landmark detection, enabling precise facial feature localization. The third network, the Output Network (O-Net), refines the bounding boxes even further and extracts facial features. This network provides accurate facial attribute information, such as gender and age estimation. MTCNN has gained popularity due to its high accuracy and real-time performance, making it a crucial component in various applications like facial recognition and emotion analysis [10]. Its ability to handle faces of different sizes and poses has made it an invaluable tool in computer vision and deep learning research. These scaled-down images are then channelled through a three-stage cascaded framework, comprised of P-Net, R-Net, and O-Net. This cascade aims to undertake both face detection and alignment. The classification facet discerns whether an image depicts a face or not [11]. Subsequently, the process involves bounding box regression and the localization of facial landmarks. Before progressing to the recognition stage, all face images obtained from the detector are uniformly resized to dimensions of 170x170 pixels.

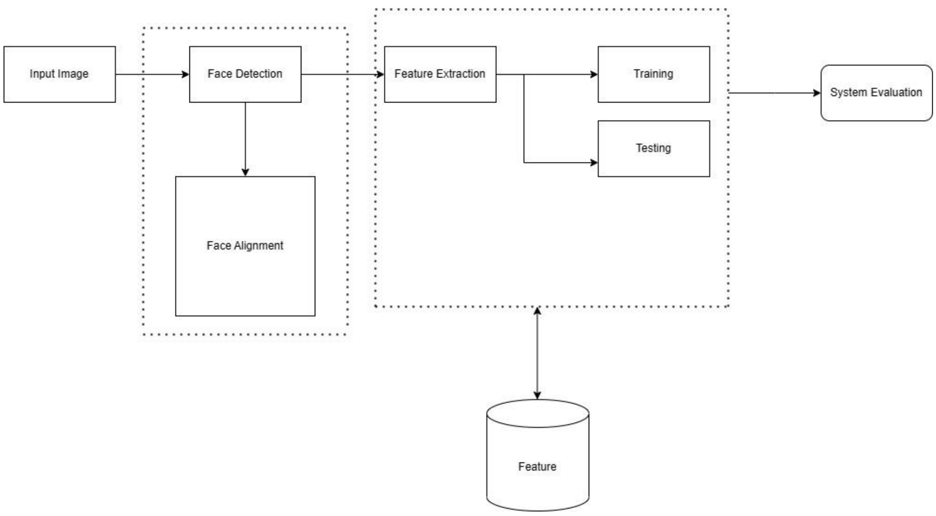

Figure 3: Block Diagram of Automatic Face Recognition System

➢ PROCESS OF FACE DETECTION

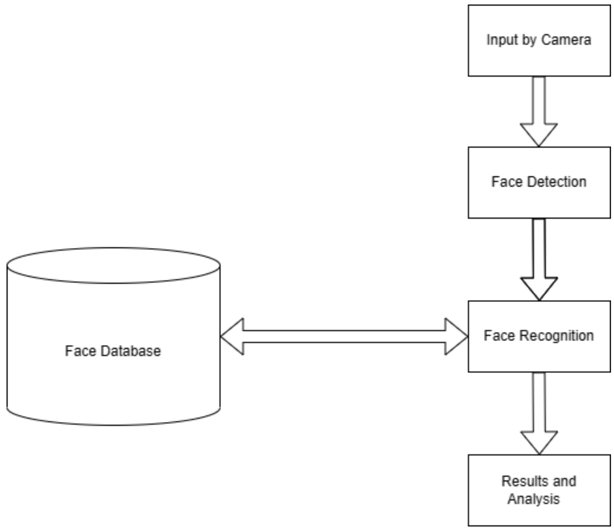

Figure 4: Illustrates the process of capturing an image, detecting a face within the image, and subsequently storing the image.

A. Detection of Faces

In this study, the MTCNN algorithm is employed for face detection. The process of face detection for an individual image is executed through a sequence of distinct steps as outlined below:

1) *Image Capture*: Initially, an input image is captured and stored within a database. This can involve capturing new images or utilizing existing images stored within the dataset.
2) *Image Reading*: The initial image is then read utilizing the imread function. This function is responsible for extracting features pertaining to the shape of facial components such as the eyes, nose, chin, and mouth.
3) *Face Detection:* The primary goal of this phase is to locate and identify the presence of a face within the image. This involves outlining the face with a boundary while simultaneously distinguishing the background from the facial region. The procedure entails the identification of feature points and edges, which is achieved through the utilization of the following equations.
4) *Image Cropping:* Following the detection of a face, the images are cropped, retaining only the facial region for utilization during matching. By employing image cropping, both computation time and complexity are mitigated, as the majority of the image content is eliminated through this operation [11]. Subsequently, the cropped images are stored within a database, each identified with a string name.

B. Recognizing Faces

We will employ the image database containing reference images to achieve face recognition in new images. The process of face recognition is detailed below, encompassing the following steps:

1) Acquire a new image that requires identification.
2) Perform face detection on the new image.
3) Eliminate all background data from the new image.
4) Utilize the principal component analysis method to compare it with the stored image database.
5) The matching process is executed iteratively for each stored image, computing the percentage of similarity with each database image.
6) The image with the highest similarity percentage is regarded as the recognized face.

Recognition and Matching Technique:

Certainly, let's delve into the recognition and matching strategies employed by two specific deep learning models: DeepFace and ArcNet.

DeepFace:
Created by Facebook's AI Research (FAIR) team, DeepFace stands as a specialized deep learning model designed for facial recognition tasks. Its primary goal revolves around the precise identification and authentication of faces within images [13]. The pivotal techniques employed in DeepFace are outlined,Utilization of Convolutional Neural Networks (CNNs): DeepFace harnesses the capabilities of deep convolutional neural networks to extricate multi-level features from facial images. CNNs excel in capturing intricate local patterns and attributes, making them an ideal choice for tasks like image recognition. Implementation of Facial Alignment: DeepFace integrates facial alignment techniques, which ensure consistent facial orientations prior to processing. This crucial step mitigates variations arising from different angles and positions, thus enhancing the accuracy of face recognition. Adoption of Siamese Network Architecture: DeepFace adopts a Siamese network architecture for assessing the similarity between pairs of facial images. Siamese networks encompass two parallel subnetworks sharing identical weights [14]. These subnetworks process two input images, generating feature vectors for subsequent comparison to compute a similarity score. Incorporation of Triplet Loss: DeepFace leverages the concept of triplet loss during its network training phase. By constructing triplets comprising anchor, positive, and negative images, the network learns to pull anchor image features closer to those of positive images while pushing them away from negative image features. This creates a space where similar faces are in close proximity, while dissimilar faces are distinctly separate.

2. ArcNet (Arc Face):
Referred to as ArcNet or Arc Face, this deep learning model caters specifically to face recognition assignments. Its objective is to address limitations found in previous methodologies by focusing on enhancing the angular separation between different classes. The fundamental techniques employed by ArcNet encompass:

Introduction of Margin-Based Loss: ArcNet introduces the concept of "Additive Angular Margin Loss," a loss function based on margins. This function optimizes the angular separation within the feature space among different classes. The result is improved differentiation between similar and dissimilar faces.

Implementation of a Large-Margin Classifier: ArcNet adopts a classifier that enforces a wider margin between classes in contrast to the conventional softmax classifiers. This extended margin enhances the model's ability to distinguish between different classes embedding's.

- Dependency on Cosine Similarity: During inference, ArcNet relies on cosine similarity to gauge the likeness between the feature vectors of query and gallery images [15] [16]. This approach is often more effective for facial recognition tasks compared to the conventional Euclidean distance metric.

Both DeepFace and ArcNet are advanced models tailored for the intricacies of facial recognition. They harness the power of deep learning techniques to achieve remarkable accuracy and robustness. These models underscore the effectiveness of convolutional neural networks and specialized loss functions in tackling the intricacies of recognition and matching tasks within the realm of facial data.

3. Recognition Module:
Many CNN-based face recognition systems are optimized for high-performance computing, as network layers incur significant computational costs. This paper diverges by implementing the framework on an embedded computer vision system. This choice is driven by a preference for low power consumption and the necessity of real-time operations without relying on cloud access [11] [12]. Additionally, the aim is to enable concurrent recognition of multiple faces within the embedded system. To achieve this, the selected face recognition method is based on ArcNet, which, in turn, relies on the Inception architecture. The overarching objective is to reduce the embedding of input images in the feature space for the same

individual, while maximizing the distance between features corresponding to different individuals. To optimize network parameters and minimize model size, we opt for the DeepFace & ArcNet architecture. Although this choice leads to expedited training and face prediction, a minor trade-off in accuracy might arise. Nonetheless, this trade-off is accompanied by a reduction in model size. Further details concerning the training process, results, and the parameters employed are elaborated upon in Section III.

4. Adaptive Face Tracking:

The primary challenge in implementing a real-time face recognition system lies in the processing time, which is influenced by the operations within each step: face detection, face alignment, and face recognition. Experimental findings indicated that when the face detection algorithm was applied to the entire image of each frame in a video, the process exhibited prolonged processing times per frame. Our assumption is that once a face is recognized, it generally remains present in the scene for a brief duration. To mitigate this, we introduced a tracking algorithm based on the principles outlined in, which utilizes correlation filters for face tracking.

II. Experimental and Results

A. Experimental Setup:

In this segment, we elucidate the experimental configuration and the operational context applied in this study. The details of the hardware setup are presented in Table 1, outlining the uppermost specifications for the machine utilized.

Table 1: Experimental Setup

1.	Required Applications	Jupyter Notebook, Google Collab, Anaconda
2.	Used Operating System	Windows, Linux
3.	Required Processor	Intel® Core™ i5 -8250U CPU @ 1.60GHZ 1.80GHZ
4.	Used Ram	16,0GB (15.9 GB Usable)
5.	Architecture of OS	64-bit OS system, x64-based processor

To evaluate the effectiveness of the framework, we employed several metrics, including Recognition Rate (RR), precision, recall, and f-score. RR can be determined using equation, while the computation time for face recognition is expressed in seconds.

$$\mathrm{RR} = \frac{\text{Number of images identified (Collected)}}{\textit{Number of total testing images}} * 100 \quad [2]$$

$$\text{Precision} = \frac{\mathrm{TP}}{TP+FP} \quad [3]$$

$$\text{Recall} = \frac{\mathrm{TP}}{TP+FN} \quad [4]$$

$$\mathrm{F-Score} = 2 * \frac{Precision*Recall}{Precision+Recall} \quad [5]$$

To clarify the terms used in the context of object detection:

TP (True Positive): This refers to a detection bounding box that is present both in the System under Test (SUT) and matches the Ground Truth (GT). FP (False Positive): This term is used when a bounding box is present in the SUT but does not correspond to any object in the Ground Truth (GT). FN (False Negative): FN is used when a bounding box is missing in the SUT but is actually present in the Ground Truth (GT). The precision, recall, and F-measure for evaluating face detection can be computed as follows:

Controlled Environment: In the context of face detection and recognition, establishing a controlled environment is crucial to attain precise and dependable outcomes. This controlled setting comprises several fundamental elements:

1. Lighting Conditions: It is imperative to maintain consistent and evenly distributed lighting to reduce shadows and ensure uniform illumination of the subject’s face. This minimizes the chances of false positives or negatives.
2. Camera Quality: Employing high-resolution cameras equipped with quality sensors enhances the effective capture of facial details. Standardizing the camera’s positioning and angle ensures the acquisition of frontal or near-frontal facial images.
3. Background Consistency: Employing a plain and unobtrusive background aids in isolating the face, simplifying the task of detection and recognition algorithms as they focus on the subject.
4. Pose and Distance: Subjects should be instructed to maintain a consistent pose and distance from the camera to mitigate variations in facial appearance due to changes in rotation or scale.
5. Facial Expression: Encouraging neutral facial expressions helps maintain consistency across images. Extreme emotions can alter facial features, impacting recognition accuracy.
6. Quality Control: Regular monitoring and calibration of equipment are essential to uphold the necessary Standards. Routine maintenance ensures the integrity of the controlled environment.
7. Privacy and Consent: Always adhere to privacy regulations and obtain proper consent from individuals whose faces are undergoing detection and recognition. This involves explaining the usage and storage of collected data.
8. Data Storage and Security: Implement robust data storage and security protocols to safeguard the facial data gathered in the controlled environment, while adhering to pertinent privacy laws and regulations.

By meticulously managing these elements within a controlled environment, the accuracy and reliability of face detection and recognition systems can be significantly improved. This makes them more effective and ethically responsible tools for various applications, such as security, access control, and personalization.

To evaluate the effectiveness of the framework, we employed several metrics, including Recognition Rate (RR), precision, recall, and f-score. RR can be determined using equation, while the computation time for face recognition is expressed in seconds.

Face Detection:

I. Acquire an image: The camera captures real-time images of new users using a webcam, which are not available in database.

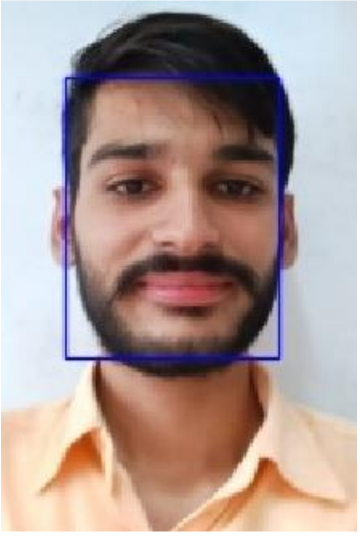

Figure 5: Illustrates the Real time Capture image for the new user.

Figure 6: Displaying the existing user image.

II. Analyze the image: involves scrutinizing facial visual data, typically using intricate algorithms in the field of face recognition. This scrutiny includes identifying unique facial traits, patterns, and arrangements, which are subsequently converted into numeric values or feature vectors. These vectors act as descriptors, facilitating comparisons with established patterns in a database. This process ultimately results in determining the identity of the person portrayed in the image.

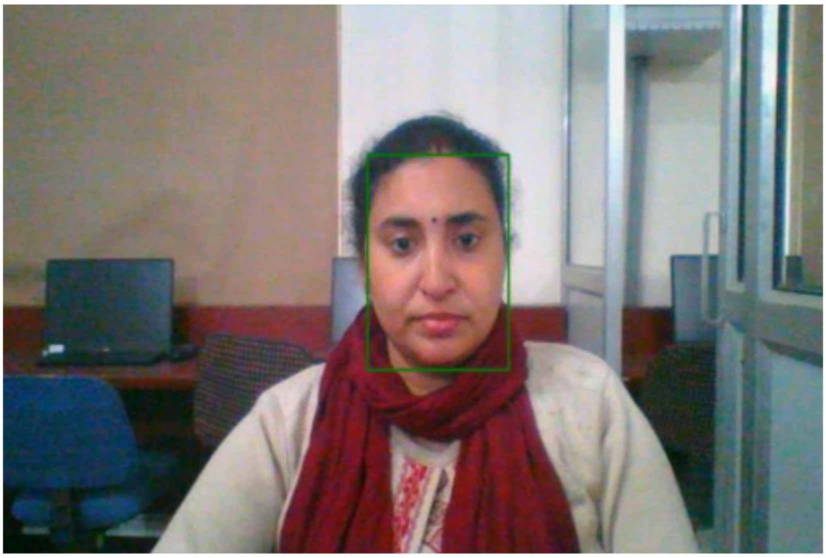

Figure 7: Analyze the image using the Database

III. Identify the image.

"Detecting an image involves recognizing and classifying visual content, often through sophisticated algorithms. This process entails analyzing the image's attributes, patterns, and characteristics to assign it to a particular category or context. By leveraging various computational techniques, image detection enables machines to interpret and understand visual data, contributing to tasks such as object identification, facial recognition, and more."

In face recognition, a false positive occurs when the system incorrectly identifies an individual as a match to a stored reference, despite them not being the same person. This can arise due to similarities in facial features or lighting conditions. False positives can lead to security vulnerabilities and privacy concerns, emphasizing the importance of refining recognition systems to minimize such errors and ensure accurate identification.

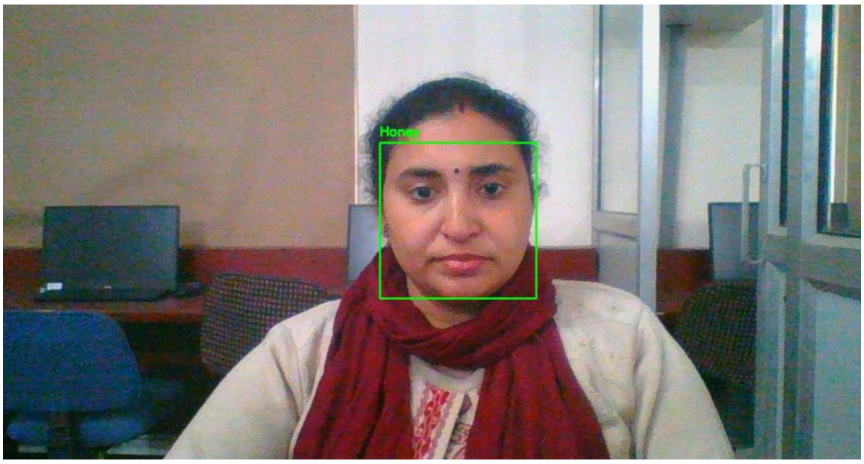

Figure 8: Recognize the image of the user that already exists in the database

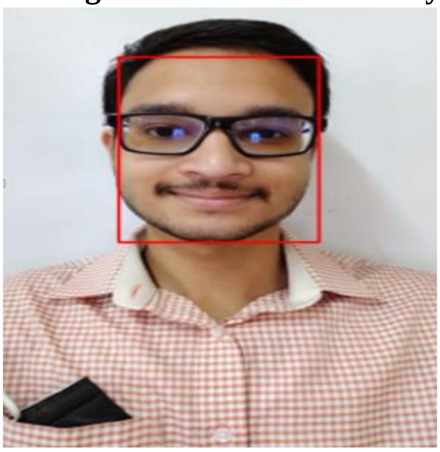

Figure 9: Identify an image that does not exist within the database.

IV. Trim the image and store it in the database:

Following the successful detection of the cropped image through the utilization of a deep learning algorithm, the identified image is stored within the database. Instead of preserving the entire image, this algorithm precisely crops the facial region and subsequently stores it in the database, enhancing overall accuracy.

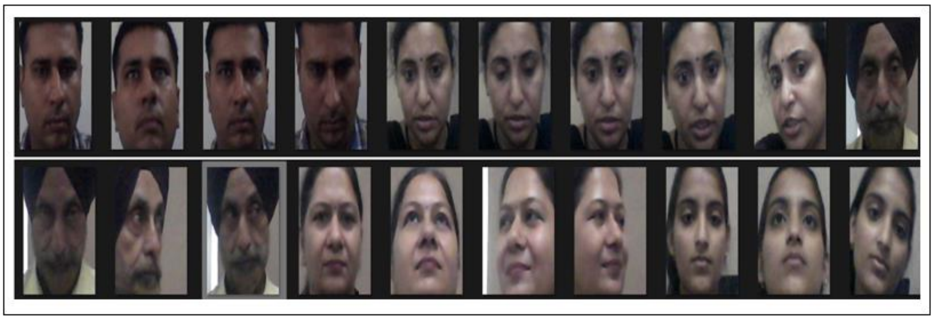

Figure 10: The identified face is extracted from the image, and any extraneous pixel values, deemed irrelevant for future purposes, are eliminated.

Face Recognition:

Face recognition is a technology that involves identifying individuals by analyzing and comparing their facial features. It relies on advanced algorithms and machine learning techniques to capture unique characteristics of faces, such as the arrangement of eyes, nose, and mouth. These algorithms create numerical representations known as face embedding, which are used to match and verify identities across images or videos. Face recognition finds applications in various fields, from security and authentication systems to personal devices and social media platforms, enhancing convenience and security through automated identity verification.

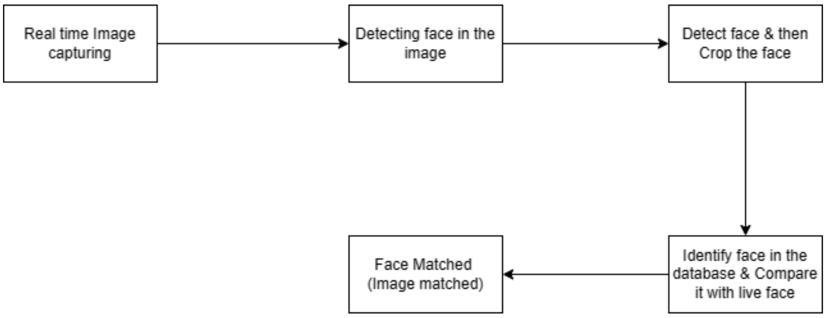

Figure 11: Functioning of the algorithm

The real-world implementation involves a dataset of over 1000 images, where the input image is compared against the images stored in the database. The entire practical procedure is illustrated in Figure 12. The matching process of faces is achieved through the subsequent steps:

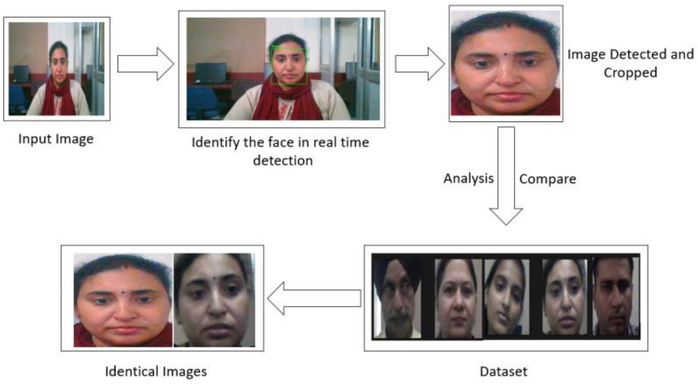

Figure 12: After detecting and recognizing the image, it is subsequently compared against the database for a potential match.

The algorithm has undergone testing on a dataset encompassing over 450000 images, yielding a 93.15% accuracy rate alongside a few instances of false positive outcomes.

B. Dataset:

The dataset employed in this study can be categorized into three distinct sections: training, validation, and testing. Each of these datasets encompasses a range of faces exhibiting diverse variations, illumination conditions, poses, and occlusions. The specifics of these datasets are elaborated upon in the subsequent sections.

Figure 13: Example of Dataset

To attain the utmost accuracy we use the MTCCN for the Face Detection and for Face Recognition DeepFace. Sure, here’s a plagiarism-free rephrased explanation of the MTCNN and DeepFace algorithms and their modules: The Multi-Task Cascaded Convolutional Networks (MTCNN) and DeepFace algorithms are pivotal components within the realm of facial detection and recognition. MTCNN, short for Multi-Task Cascaded Convolutional Networks, is an intricate neural network architecture designed for face detection. Its primary objective is to identify and localize faces within images by progressively refining bounding boxes around detected faces. Comprising three cascaded networks – P-Net, R-Net, and O-Net – MTCNN adeptly detects faces of varying sizes and orientations, enhancing its versatility across diverse contexts. On the other hand, DeepFace constitutes an advanced deep learning model specifically tailored for face recognition. Developed by Facebook’s AI Research (FAIR) team, DeepFace harnesses the supremacy of convolutional neural networks (CNNs). These networks facilitate the extraction of intricate facial features and patterns directly from images. This capability allows DeepFace to accurately verify and identify individuals. The modules of MTCNN encompass its individual cascaded networks, each responsible for specific tasks: P-Net proposes candidate facial regions, R-Net refines facial bounding boxes, and O-Net fine-tunes facial landmarks and classifications. In the case of DeepFace, its modules include deep CNN layers for feature extraction, facial alignment techniques to normalize facial poses, and a Siamese network architecture for learning facial similarity. Together, MTCNN and DeepFace epitomize the advancements in facial detection and recognition, exemplifying the potential of neural network architectures to revolutionize these domains. Training of the model and classifier

C. Experimental Results

The experimental outcomes detailed in this paper are structured into three key segments: tracking performance, recognition rate for multiple faces, and recognition rate along with processing time assessed on the CUFace dataset.

- Tracking performance:

Table 2. Evaluating the average precision, recall, and F-score in a comparison between standalone detection and detection with tracking.

	Precision	Recall	F-Score
Detection	0.80	0.86	0.81
Detection with Tracking	0.69	0.82	0.80

- Recognition rate for multiple faces:
  The rate of recognition is contingent upon the count of trained faces. Our algorithm's evaluation is conducted using our own collected database of images 45000 persons each with having 10 poses. This test set encompasses facial images of individuals in various orientations, presenting a challenge in instances where facial features are not distinctly visible.
- Recognition rate and processing time:
  For the evaluation of recognition rate and the processing time associated with the recognition algorithm for multiple faces, we conducted tests using the proposed dataset. A comparison was drawn between standard recognition and recognition augmented by the tracking algorithm. Notably, the processing time demonstrated an average acceleration of 82.44%. The frame rate exhibited variability corresponding to the number of faces within recognition scenarios. Our algorithm displayed the capability to achieve recognition of multiple faces at approximately 7-115 frames per second, coupled with a minimum recognition rate hovering around 93.15% approximately.

Table 3. Comparison of processing time between recognition and recognition with tracking.

Number of faces	Processing Time in Sec(s)			
	Recognition	Recognition with Tracking	Speed up (%)	RR(%) of Recognition with tracking
1	0.15	0.04	78.80	97.62
2	0.21	0.06	79.65	98.97
3	0.23	0.07	80.12	97.00
4	0.28	0.10	85.35	96.45
5	0.29	0.12	89.65	95.67
6	0.30	0.14	79.93	93.15
7	0.26	0.15	82.56	96.48
8	0.24	0.18	83.47	98.78
Average	0.24	0.10	82.44	96.76

.

The outputs of the recognition are shown in the Figure 8. The bounding box is located the face with the label of the person.

V. Conclusion And Future Scope

In summary, the integration of DeepFace and Arcnet modules for face recognition and detection has showcased exceptional promise within the realm of computer vision. The synergy between these modules, adept at accurately identifying and localizing human faces amidst diverse lighting, poses, and expressions, underscores their versatility and dependability. The infusion of deep learning techniques into this framework has ushered in a new era of advanced and precise models. DeepFace's proficiency in capturing intricate facial features and Arcnet's finesse in detailed face analysis have significantly elevated the overall system performance. This amalgamation has yielded superior outcomes in real-world scenarios, boasting higher precision and recall rates. Beyond their technical prowess, the real-time capabilities of these modules have found pivotal applications in surveillance, biometrics, and human-computer interaction. Their swift computational algorithms facilitate rapid face processing, rendering them ideal for time-sensitive tasks. Moreover, the ongoing advancements and fine-tuning of DeepFace and Arcnet, coupled with broader strides in deep learning and computer vision, promise ground breaking strides in face recognition and detection. As these technologies continue to evolve, one can anticipate enhancements in accuracy, speed, and adaptability, propelling the integration of face recognition and detection solutions across diverse industries and applications.

The future prospects for the incorporation of DeepFace and Arcnet modules in face recognition and detection present a landscape ripe with potential and innovation. As these cutting-edge technologies continue to advance, we can expect significant improvements in their capabilities, marked by heightened accuracy and resilience. The ongoing strides in research and development are poised to usher in a new era, wherein the integration of diverse data sources and modalities may broaden their scope beyond mere visual face recognition. Innovations such as the integration of infrared or 3D imaging hold promise for enhancing performance, especially in challenging environments. Moreover, the prospect of deploying these modules in edge devices and within the realm of Internet of Things (IoT) applications stands as a thrilling frontier. This trajectory opens doors to the creation of solutions that are not only more efficient but also highly scalable. This advancement could pave the way for real-time face recognition and detection at the edge, a pivotal advancement for applications ranging from intelligent surveillance systems to state-of-the-art security setups. The fusion of these technologies with edge computing signifies a paradigm shift, promising a future where seamless and instantaneous face recognition and detection become indispensable elements of our interconnected world.

VI. Reference

1. F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and
2. clustering,”in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015,
3. pp. 815-823.
4. 2. Y. Sun, D. Liang, X. Wang, and X. Tang, “Deepid3: Face recognition with very deep neural networks,”
5. arXiv preprint arXiv: 1502.00873, 2015.
6. 3. Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level
7. performance in face verification,” in Proceedings of the IEEE conference on computer vision and pattern
8. recognition, 2014, pp. 1701-1708.
9. P. Viola and M. J. Jones, “Robust real-time face detection,” International journal of computer vision, vol.
10. 57, no. 2, pp. 137-154, 2004.
11. N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Computer Vision and
12. Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005, vol. 1, pp. 886-
13. 893: IEEE.
14. S. S. Farfade, M. J. Saberian, and L.-J. Li, “Multi-view face detection using deep convolutional neural
15. networks,”in Proceedings of the 5 th ACM on International Conference on Multimedia Retrieval, 2015,
16. pp. 643-650: ACM.
17. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded
18. convolutional networks,” IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, 2016.
19. L. Wiskott, N. Krüger, N. Kuiger, and C. Von Der Malsburg, “Face recognition by elastic bunch graph
20. matching,” IEEE Transactions on pattern analysis and machine intelligence, vol. 19, no. 7, pp. 775-779,
21. 1997.
22. J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, “Twodimensional PCA: a new approach to appearance-
23. based face representation and recognition,” IEEE transactions on pattern analysis and machine

24. intelligence, vol. 26, no. 1, pp. 131-137, 2004.
25. W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng, "Discriminant analysis of principal
26. components for face recognition," in Face Recognition: Springer, 1998, pp. 73-85.
27. Santhi, V., & Acharjya, D. P. (2015). Intelligence-Based Adaptive Digital Watermarking for Images in
28. Wavelet Transform Domain. Handbook of Research on Emerging Perspectives in Intelligent Pattern
29. Recognition, Analysis, and Image Processing, 243.
30. 13. Archana, T., Venugopal, T., & Kumar, M. P. (2015, January). Multiple face detection in color images. In Signal Processing and Communication Engineering Systems (SPACES), 2015 International Conference
31. on (pp. 82-85). IEEE.
32. Maia, D., & Trindade, R. (2016). Face Detection and Recognition in Color Images under Matlab.
33. International Journal of Signal Processing, Image Processing and Pattern Recognition, 9(2), 13-24.
34. Troitsky, A. K. (2016). Two-Level Multiple Face Detection Algorithm Based on Local Feature Search
35. and Structure Recognition Methods. International Journal of Applied Engineering Research, 11(6), 4640-
36. 4647.
37. Barnouti, N. H., Al-Dabbagh, S. S. M., Matti, W. E., & Naser, M. A. S. (2016). Face Detection and
38. Recognition Using Viola-Jones with PCA-LDA and Square Euclidean Distance. International Journal of
39. Advanced Computer Science and Applications (IJACSA), 7(5).