Versatile Face Recognition and Tracking System

P Rajeshr[1], K. Sai Yashwitha[2], K. Lakshmi Narasimha Naidu[3],
K. Tharun Kumar[4], G. Jagadish[5]
Seshadri Rao
Assistant professor, UG Scholar[2,3,4,5]
Department of Computer Science and Engineering
Gudlavalleru Engineering College,
Gudlavalleru, Andhra Pradesh, India -521356

ABSTRACT

The Versatile Face Recognition and Tracking System is a Multifaceted tool for precise and efficient face recognition across various applications. It uses Multi-Task Cascaded Convolutional Neural Networks (MTCNN) for accurate face detection, followed by Face Net and an SVM classifier for robust face recognition. The system also incorporates YOLOv8 Tracking for real-time and effective tracking, enhancing the tracking process through advanced object detection techniques. This system is suitable for security, surveillance, and human-computer interaction, ranging from access control in secure facilities to crowd monitoring in public spaces. Its accuracy and adaptability make it a new era of sophisticated face recognition and tracking capabilities.

Keywords: Face Detection, Real time Face tracking, Face Recognition, MTCNN, FaceNet, YOLOv8, SVM.

INTRODUCTION

The need for reliable and effective face recognition and tracking systems has grown in importance and is now essential in many different fields in the age of modern technology.

The "Versatile Face Recognition and Tracking System," a creative approach that could revolutionize the field of face recognition technology, is presented in this paper. Our approach is based on the convergence of deep learning, computer vision, and artificial intelligence, which addresses the complex problems of monitoring and identifying faces in a variety of contexts. The widespread use of surveillance systems, along with the growing need for intelligent human-computer interaction and secure access management, highlights the vital significance of sophisticated facial recognition technologies. Conventional techniques, although effective in controlled settings, are frequently unable to handle the complexities presented by changes in lighting, dynamic scenarios, and facial expressions.

Our system uses state-of-the-art techniques to address these issues and offer a thorough hand flexible solution. Multi- Task Cascaded Convolutional Neural Networks (MTCNN), a powerful technique for accurate face detection, are the foundation of our system. Accurate localization and extraction of facial areas are made possible by MTCNN's ability to capture complex facial features. This is a crucial step before the next ones, in which Face Net works with a Support Vector Machine (SVM) classifier to guarantee reliable face identification. Face Net, which is well known for mapping facial traits into a high-dimensional space, makes it easier to recognize faces accurately from a variety of people, regardless of lighting or position differences. At the same time, our technology recognizes that real-time face tracking is necessary in dynamic contexts and easily integrates YOLOv8 Tracking. Modern object identification framework YOLOv8 (You Only Look Once) improves our system's performance by offering effective real-time face tracking. Its capacity to process images in a single forward pass, eliminating the need for intricate post-processing, is in line with the changing needs of applications like surveillance and crowd monitoring.

With MTCNN, Face Net, SVM, and YOLOv8, our system is position edasa flexible instrument that can handle a wide range of tasks. Our system's versatility makes it a dependable and essential tool, whether it issued in public areas where continuous and efficient crowd monitoring is required or in controlled entry points with strict security requirements.

This paper offers a thorough analysis of the system's methods, experimental findings, and a thorough discussion of the system's possible applications in the fields of security, surveillance, and human-computer interaction. This Versatile Face Recognition and Tracking System ushers in a new era of advanced facial recognition capabilities by setting a new benchmark in accuracy and adaptability. All things considered, the subject of this study stands out as an innovative pioneer in the rapidly changing field of face recognition technology. Our solution anticipates the needs of dynamic and intelligent environments by combining cutting-edge algorithms, which not only satisfies present demands for accuracy and adaptability but also anticipates future ones. The following sections of this paper will explain the nuances of our approaches, showcase the empirical findings, and have a critical conversation about the wider consequences andu ses of our system.

RELATED WORK

Recent advances in artificial intelligence, deep learning, and computer vision have led to amazing gains in face identification and tracking systems. An outline of the current environment is given in this section, with special attention to important approaches and innovations that have made it possible to construct a Flexible Face Recognition and Tracking system.

With the help of numerous researchers and methodological advancements, the field of face recognition technology has changed over time. Ning Zhang, Junmin Luo, and Wuqi Gao (2010) established the foundation for face detection technology based on Multi-task Cascaded Convolutional Networks (MTCNN), which paved the way for later developments and served as a fundamental resource for researchers pursuing the field.

In 2016, Yu Qiao, Zhanpeng Zhang, Zhifeng Li, and Kaipeng Zhang presented a groundbreaking idea: joint face detection and alignment via MTCNN. This work represented a significant advancement, highlighting the integration office detection and alignmenttasks—a crucial component of reliable facial recognition systems.

A facial recognition-based surveillance system was presented by Edwin Jose, Greeshma M., Mithun T. P. Haridas,andM.H.Supriyaintheyear2019.Their study,which demonstrated the actual deployment of FaceNet and MTCNN on Jetson TX2, contributed to the practical use of facial recognition in surveillance.

2020 witnessed a number of noteworthy contributions from various angles. Dong Cui, Qi Guo, Zhihui Wang, and Caixia Wang investigated a multi-face identification method used in video surveillance. Mahmudul Hasan Robin, Abu Mohammad Taief, Md. Minhaz Ur Rahman, Qamrun Nahar Eity, and Mahmudul Hasan Robin worked simultaneously to improve face and eye detection performance using MTCNN. By addressing issues unique to surveillance contexts, these studies improved our knowledge of face recognition in dynamic environments.

2020 saw the development of a Face Recognition System using the Facenet algorithm by Ferry Cahyono, WirawanWirawan,andReza Fuad Rachmadi, withaspecific application to employee presence. This demonstrates the various uses of facial recognition technology in attendance monitoring and labor management.

Following the chronological order, Yufei Bao and Rong Dangtackled the particular problem office detection in non-uniform low light in 2021. Through their effort, the MTCNN framework was enhanced, and face detection was successfully adapted to difficult lighting conditions.

In 2021, T. R. Ganesh Babu, K. Shenbagadevi, V. Sri Shoba, S. Shrinidhi, J. Sabitha, and U. Saravanakumar explored image processing methods for face recognitionusing machine learning techniques. This study provided insights into the broader landscape of image processing techniques complementing face recognition methodologies.

WeijiaFeng,SiyaoQi,XinyuZuo,andI.G. Naveen made the most recent contribution in 2022. They present a modern combination of these two key technologies for improved face recognition accuracy with their work, which uses a Face Recognition Model based on MTCNN and Face net.

Face recognition underwent a radical change with the introduction of deep learning. Convolutional neural networks, or CNNs, have become highly effective instruments for extracting features, which has made it possible to create reliable and accurate recognition models. Although the literature now in publication offers awiderange of approaches, a complete system that smoothly combines precise face recognition with real-time monitoring is still up for grabs. By fusing the best features of MTCNN, Face Net, SVM,andYOLOv8,the"VersatileFaceRecognitionandTracking System" seeks to close this gap and establish new benchmarks for accuracy and adaptability.

DATASET DESCRIPTION

Our data is arranged into two main subfolders in the dataset directory: "train" and "test." The data used to train our model isstored in the"train" folder, while the data used to test and assess the model's performance is stored in the "test" folder. The subfolders "person_1" through "person_50," which each identify a distinct dataset participant, are located inside the "train" and "test" folders.

All of these individual subfolders act as storage for an assortment of images. Each individual has been assigned eight photos (image1.jpg through image8.jpg) in the training set. This composition makes up eighty percent of the training dataset.

Now that we have reached the testing set, which makes up 20% of the total dataset, we have added two more images (image9.jpg and image10.jpg) for each person. The purpose of this addition is to include a somewhat larger set of photographs for examination in order to create a more difficult test scenario.

Our dataset consists of 500 images altogether, carefully split between training and testing sets. 80% of every person's photos are used in the model's training process thanks to the thoughtful distribution, which promotes efficient learning and pattern recognition.

The remaining 20%, meanwhile, creates a unique set that isused to evaluate the model's performance and generalization on untested data.

This configuration makes it easier to train validate, and assess models on different data subsets. By using this methodical approach, we want to improve the model's performance and adaptability in a variety of settings, which will ultimately help our Versatile Face Recognition and Tracking System succeed.

PROPOSEDSYSTEM

The proposed system encompasses a comprehensive pipeline that seamlessly integrates advanced face recognition and tracking technologies. This section elaborate sonthemeticulous process of data preparation the intricate steps involved in building the face recognition model, and the real-time video testing capabilities of the system.

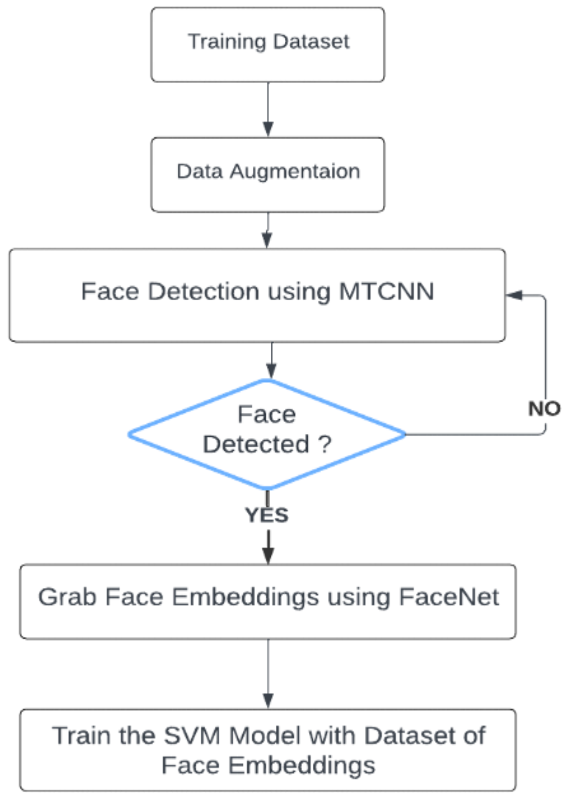


Fig2: Training Work flow

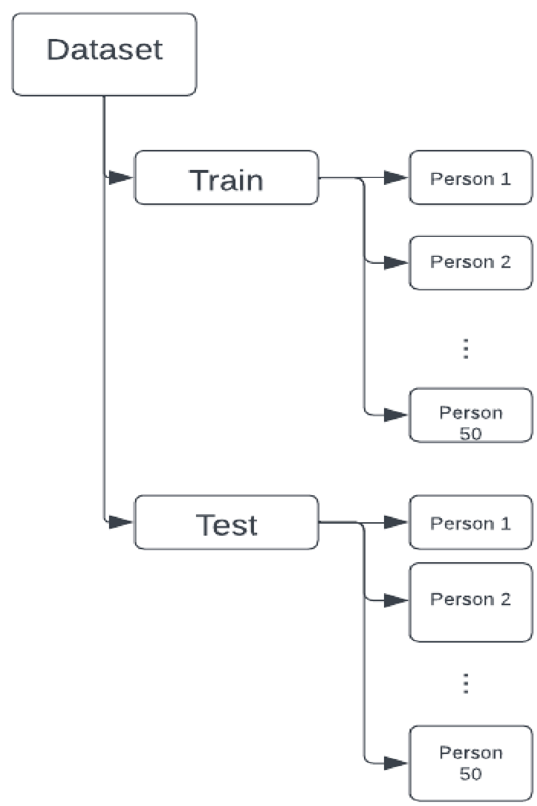


Fig1: Dataset Directory

1.1. Data Augmentation

Before employing the MTCNN framework for face detection, we applied data augmentation techniques to improve the quality and variety of the input images. This preprocessing step included rotating, scaling, and horizontally flipping the images to guarantee the model could detect faces in a variety of orientations and conditions. Additionally, random brightness and contrast adjustments were applied to simulate varying lighting conditions in thereal world. The augmented data set offered a more extensive training set for MTCNN, enhancing its resilience and accuracy in detecting faces in a variety of situations. This dualstrategy, combining data augmentation at both the face detection and Face Net stages, mutually enhanced the efficaciousness of our flexible face detection system.Face Localization

The precise and effective identification of faces inside picture or video frames forms the basis of our face recognition and tracking technology. We use the Multi- Task Cascaded Convolutional Neural Network (MTCNN), a cutting-edge deep learning model created especially for face detection problems, to do this. MTCNN functions in three stages: face localization, landmark point detection, and bounding box refining. Each step focuses on a distinct face detection task. The network finds possible face regions in the first step and creates bounding boxes to surround these regions. The eyes, nose, mouth, and other facial markers are carefully located in the second stage, which refines these bounding boxes. In order to ensure that the detected face is optimally covered, MTCNN further refines the bounding boxes in the third step.The Multi-Task Cascaded Convolutional Neural Network (MTCNN) face identification method is essential to the operation of our system. We import the carefully produced dataset, which includes training and testing images, and then we incorporate MTCNN, a deep learning model that is specifically designed for face detection applications. MTCNN strategically finds and refines possible face areas in pictures or video frames by using its pre-trained weights in three different steps. In order to achieve this, bounding boxes around faces that are identified must be created, facial landmarks must be found for accurate placement, and bounding boxes must be adjusted for maximum coverage. A crucial stage in further processing is the extraction of face pixels from each image or frame, which is guided by the ensuing coordinates of these revised bounding boxes. With a distinct division between the training and testing datasets, these extracted face pixels are arranged and saved in arrays.In order to optimize organizational effectiveness, the face pixels are assembled into a NumPy zip (NPZ) file for easy access in later system phases. Our face identification and tracking solution is based on this careful application of MTCNN, which guarantees reliable and accurate face detection for further processing.

1.2. Face Embeddings

Once MTCNN has successfully recognized faces,the next critical step in our face recognition system is to convert the discovered face pixels into compact and meaningful representations called face embeddings.

We use the FaceNet model, a pre-trained deep learning model created especially for producing high-dimensional embeddings that capture the unique characteristics of each face, to achieve this goal. Following their acquisition by MTCNN, the face pixels are supplied into the Face Net model, which functions by projecting faces into a high-dimensional space.

It is essential to employ a pre-trained Fac eNet model since it has been trained on large datasets to acquire features that are generalizable and skilled atc atching the subtleties of different facial emotions, positions, and lighting situations.

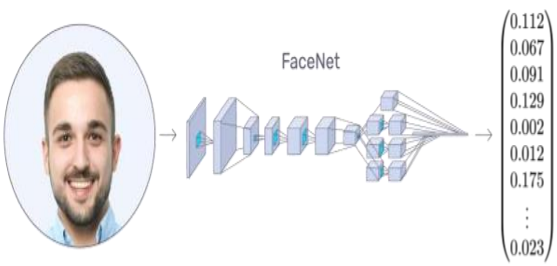
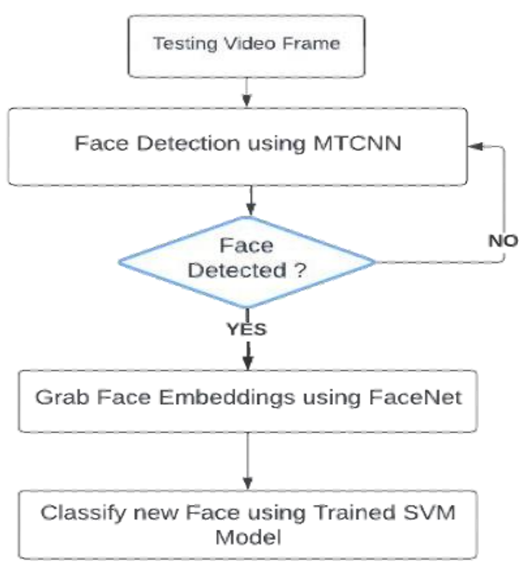


Fig3: Face net Embeddings

A numerical vector or embedding specific to each face is FaceNet's output. The training and testing datasets provide these embeddings for every face, giving rise to a condensed but rich in information representation of facial traits. Then, for organizational clarity, these embedding sare



kept separate from testing embeddings by being kept in files. Robust face recognition is based on the face embeddings that are acquired by FaceNet. The system is able to identify and measure minute variations between faces by converting facial data into a mathematical representation. This is crucial for the pipeline's laterphases, particularly for training the Support Vector Machine (SVM) classifier to recognize faces accurately.

After obtaining useful face embeddings from the Face Net model, we use the seembeddings to train a Support Vector Machine (SVM) classifier in the next step of our face recognition pipeline. Strong in pattern recognition and prediction based on learnt features, the SVM classifier is a machine learning method. The SVM classifier in our system is trained on a comprehensive set of training face embeddings, which enables it to effectively and reliably categorize faces in real-time testing. The labeled face embeddings from the training dataset are fed into the SVM classifier during the training process. The identification of the person each embedding represents is linked to it, resulting in a supervised learning situation. The SVM gains the ability to efficiently separate the embeddings of several individuals by constructing an optimal decision boundary in the high-dimensional embedding space

EXPERIMENTS AND RESULTS

The face embeddings from the testing dataset are used to thoroughly assess the SVM model's effectiveness once it has been trained. The system evaluates its capacity to reliably identify faces and classify them by feeding testing embeddings into the trained classifier. By employingpertinentevaluationcriteriatogaugethec lassifier'saccuracy, its dependability in practical situations is guaranteed. We have tested accuracymetricsont he test data. And thetested model is also applicable for real-time scenarios too.

5.1 . Testing using the test data

Fig4: Testing Work flow

When the trained SVM classifier reaches a

high enough accuracy level, it is exported and stored as a ".sav" file. The learned patterns and decision limits are contained in this file, which makes it possible to integrate them easily into our system's real-time video testing component. The finished product of the training step is the exported model, which is prepared for use in dynamic contexts for precise and efficient face recognition.

The datasets we taken for testing the trained SVM model contains images of several people faces that are already seen by the model. There were few faces where the model has to mreport as unknown rather than mapping them to someone. We have tested the model and we obtained training accuracy as98% and testing accuracy 90%.

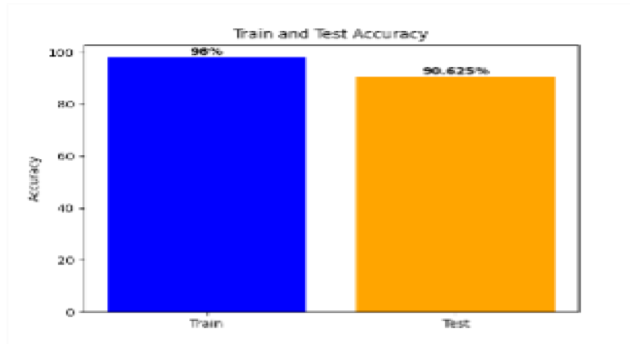


Fig5: Training and Testing Accuracy.

5.1.Testing using Real world Scenarios

Upon switching to real-time video testing, the learned SVM classifier becomes an essential part of the system. Following processing of each frame, the identified identity is shown on the frame after the embedded face is classified using the trained model. With this connection, the system is guaranteed to be quick and precise, which makes it useful in a variety of real-life situations.

Fig6 :Face Recognition

we utilized testing datasets for face recognition, employing MTCNN for face localization and FaceNet for extracting face embeddings. The SVM-trained classifier effectively classified the faces based on these embeddings. In real-time scenarios, the system processed video frames by applying the same procedures, demonstrating the adaptability of our approach for dynamic environments and achieving accurate face recognition.

Tracking of the detected face is done using YOLOv8 tracking. Its architecture allows simultaneous detection and tracking by dividing the image into a grid and predicting bounding boxes and class probabilities foreach grid cell. The integration of YOLOv8 tracking enhances our system's capability to robustly follow and identify faces in dynamic and challenging scenarios.

CONCLUSION

Our study presents the "Versatile Face Recognition and Tracking System" as a ground-breaking solution that establishes new standards in the area, in the never-ending quest of accuracy and flexibility in face identification and tracking systems. Our system combines state-of-the-art technologies to provide a flexible and reliable platform, starting from the careful steps of data preparation, which include using a variety of datasets for training and testing, to the complex phases of face identification, embedding, classification, and real-time tracking. Using Multi-Task Cascaded Convolutional Neural Networks (MTCNN) earlyon guarantees precise andeffective face identification, which sets the stage for later steps. Nuanced recognition is made possible bythe FaceNet model's conversion of identified face pixels into high-dimensional embeddings, which captures the substance of facial features. By including a Support Vector Machine (SVM) classifier that has been trained on these embeddings, a strong face recognition model that can accurately identify faces even in dynamic contexts is produced. The system's usefulness is increased by real-time video testing, and YOLOv8's object detection and tracking capabilities add a fresh perspective. Every person is given a distinct tracking ID, which guarantees ongoing observation across frames. Face recognition and YOLOv8 tracking work together seamlessly to produce a comprehensive and flexible system that has the potential to transform security, surveillance, and human-computer interaction applications.

The importance of this system become sclear when we consider the path from data preparation to real- time tracking. With the efficiency of YOLOv8 tracking and flexibility to a wide range of facial traits and expressions, our system is at the forefront of advanced face recognition technology. Its uses range from dynamic crowd monitoring to secure access control, bringing in a new era of intelligent surroundings.

In conclusion, this study presents a "Versatile Face Recognition and Tracking System" that anticipates the changing requirements of dynamic situations in addition to addressing the current issues in facial recognition. Our solution not only demonstrates the current state of face recognition technology but also pushes the field toward new frontiers in accuracy, adaptability, and practical application by seamlessly integrating cutting-edge technologies. The progression from data preparation to real-time tracking represents a fundamental change in our understanding of and approach to face recognition and tracking technology, laying the groundwork for a time when intelligence and flexibility will naturally come together.

Fig 7: Face Recognition and Tracking

FUTURE SCOPE

The "Versatile Face Recognition and Tracking System" has a lot of exciting potential for development and growth in the future. We want to increase accuracy by adding a wider variety of circumstances and phrases to the dataset. Face identification and recognition will be improved even more by investigating state-of-the-art deep learning techniques. By bringing on-device capabilities through integration with edge computing systems, dependency on centralized resources may be decreased. For a comprehensive system, we also anticipate possible partnerships with cutting-edge technology such as face emotion detection and privacy-preserving methods. Our continuous efforts will prioritize addressing ethical concerns and enhancing real-time tracking in dynamic contexts. In order to satisfy the changing needs of a dynamic technological world, the system will remain at the forefront of innovation through ongoing model upgrades and multidisciplinary partnerships.

REFERENCES

1. Ning Zhang, Junmin Luo, Wuqi Gao, “Research on Face Detection Technology Based on MTCNN,” 2010.
2. Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Yu Qiao, “Joint Face Detection and Alignment using Multi-task Cascaded Convolution Networks” 2016.
3. Yufei Bao, Rong Dang, “Face Detection under Non- uniform Low Light Based on Improved MTCNN” 2021.
4. Qi Guo, Zhihui Wang, Caixia Wang, Dong Cui, “Multi-face detection algorithm sutiable for video surveillance” 2020.
5. Mahmudul Hasan Robin; Md. Minhaz Ur Rahman; Abu Mohammad Taief; Qamrun Nahar Eity, “Improvement of Face and Eye Detection Performance by Using Multi-task Cascaded Convolutional Networks”, 2020.
6. T.R. Ganesh Babu; K. Shenbaga devi; V.Sri Shoba;S. Shrinidhi;J. Sabitha;U. Saravana kumar,” Image Processing Methods for Face Recognition using Machine Learning Techniques”, 2021.
7. Ferry Cahyono; Wirawan Wirawan; RezaFuad Rachmadi, “Face Recognition System using Face net Algorithm for Employee Presence”, 2020.
8. Siyao Qi; Xinyu Zuo;Weijia Feng;IG Naveen, “Face Recognition Model Based On MTCNN and Face net” 2022.
9. Edwin Jose; Greeshma M.;Mithun T.P. Haridas; M.H. Supriya “Face Recognition based Surveillance System Using Face Net and MTCNN on Jetson TX2 ,” 2019.
10. Luka Dulcic “Face Recognition with FaceNetandMTCNN”2019.