Face Recognition in Dense Crowds Using Deep Learning from IP Camera

Dr M Sobhana, Mudunuri V CH S Raju, Bommaganti S V B Vikas, Kalle Bhavya Vani
[1]Department of Computer Science and Engineering, V R Siddhartha Engineering College
[2,3,4]Department of Artificial Intelligence and Data Science, V R Siddhartha Engineering College
[1]sobhana@vrsiddhartha.ac.in, [2]satyaramaraju1234@gmail.com, [3]vikas2004bommaganti@gmail.com, [4]bhavyavanikalle@gmail.com

Abstract

The facial recognition system is a way to maintain security, authentication, and high-level surveillance of individuals in a crowded area. The traditional way of identifying and monitoring individuals in a congested area fails one or the other time. So, to improve security there is a need for a model that automatically detects and recognizes the faces of individuals. The proposed model is an advanced way to detect and recognize individuals by their names or as unknowns in a crowded area. Face detection is the primary stage where MTCNN (Multi-Task Cascaded Convolutional Neural Network) is used, it accurately locates faces and generates bounding boxes around the face as its outcome, these detected faces are been recognized by FaceNet. The result can be either with the respective names of the individual and the other as unknown after detection and recognition. Model achieved an accuracy of 95% on the test set. This strategy of surveillance ensures security and streamlined access control.

Keywords. Facial Feature Detection, MTCNN, Face recognition, FaceNet

1. INTRODUCTION

Biometrics and surveillance are the basic ways of maintaining authentication and security which are made must in every workplace and crowded area. In real-time scenarios due to huge crowds in every corner of the world, the traditional way of surveillance is a risk-including task. Apart from this, there is a significant challenge where people are occluded by others. Considering these challenges, we require a model that detects and recognizes individuals. This proposed solution is a DL-based face recognition model that is trained to identify individuals regardless of occlusion. Facial recognition system maintains security, and surveillance regardless of crowd, occlusion.Recognizing faces accurately depends on the dataset used while training the model. This data collection phase is the beginning stage which includes capturing a variety of faces of individuals, both normal and occluded faces. The dataset consists of a minimum of 200 images per individual to ensure covering different angles and occlusion patterns. As the dataset includes several images of an individual this helps the model to identify faces under any circumstances and makes it work efficiently than the traditional face recognition system. The proposed model uses deep learning techniques, such as MTCNN[1]to perform face detection. This technique uses a cascaded structure to locate faces in the images accurately under complex conditions like occlusions, varied poses, and lighting changes. Its high frequency and speed make it suitable for surveillance and security in dense crowds. FaceNet is a technique used in the face recognition phase which comes next after the detection process. The detected faces in the video stream or the image are passed to FaceNet[1] for face recognition and face verification tasks. This approach ensures the model can handle a wide range of faces and any crowded conditions effectively. The process results in the identified individual in the video stream or the image with bounding boxes and facial landmarks along with their respective names or as unknown if their face is new to the model. This provides authentication of individuals in crowded and congested areas

2. RELATED WORKS

Gurlove Singh et al. [2] explored face recognition, a critical element of individual authentication. Face detection and recognition are the phases. The work examined the Eigenface and Fisher face methods, focusing on digital image processing. It identified accuracy limitations, particularly in frontal view detection due to limited adaptability to scale and rotation. The work suggests integrating an eye detection system to improve performance. This system holds promise in surveillance, mugshot matching, and potential applications in ATM and home security systems, with anticipated advancements in computer vision.

Edeh Michael Onyema et al. [3] conducted a study on improving facial expression recognition using a Convolutional Neural Network (ConvNet). They used the FER2013 dataset for training, containing seven facial expressions. Their approach significantly enhanced recognition accuracy. The study addresses the computational difficulties in facial expression recognition and underscores the potential of deep. Learning for applications in healthcare and human-machine interfaces. Their computationally efficient model can be integrated with other systems, enhancing facial recognition accuracy.

Ye Ming et al. [4] address limitations in existing facial expression recognition algorithms, focusing on the incorporation of attention mechanisms. CNN-ALSTM method is developed on the foundation of convolutional neural network and long short-term memory (CNN-LSTM), enhancing important region information mining. Additionally, a two-layer attention mechanism (ACNNA-LSTM) is introduced. Fer2013 and processed CK+ datasets are used for comparative study indicating the superiority of ACNN-ALSTM over existing methods. The study further investigates the impact of network depth and hidden layer nodes in LSTM models, highlighting their positive influence on recognition accuracy.

Mohsen Heidari et al. [5] employed a siamese network approach with transfer learning from a VGG-16 model for face recognition. Two similar CNNs process pairs of face images, utilizing a similarity criterion to determine if they belong to the same person. By using the LFW dataset the rate of accuracy is 95.62%. To advance this research, explore alternative CNN architectures within the Siamese network, capable of extracting both high and low-level features. Furthermore, adopting the "triplet loss" method and employing data augmentation techniques may enhance performance, particularly for small-sample datasets.

3. PROPOSED METHODOLOGY

The approach follows a well-structured pipeline combining deep feature extraction learning with traditional machine learning for classification as in Figure 1. It starts with collecting a diverse set of facial images, followed by preprocessing to enhance image quality through resizing and normalization. Next, a face detection model identifies facial regions, which a deep learning model then processes to generate feature embeddings. These embeddings, which serve as unique facial representations, are then classified using a Support Vector Machine (SVM)[6]. This method ensures a balance between speed and accuracy, making the system both efficient and reliable.

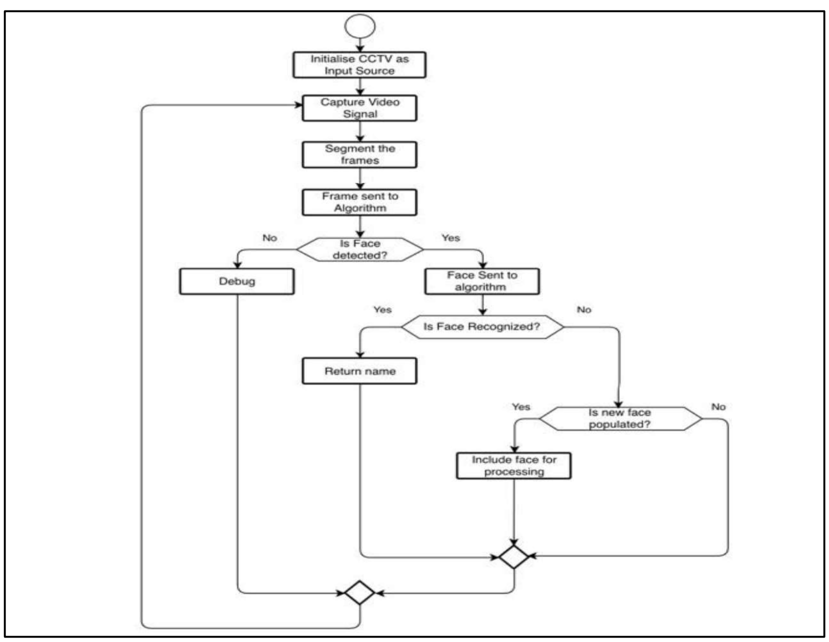

Figure 1: Proposed Methodology

3.1. Data Collection

For this study, we curated a dataset comprising facial images sourced from various public and private datasets. Each image is labeled according to the individual's identity, enabling supervised learning. To improve the model's generalizability, we included diverse variations such as changes in illumination, pose, and facial expressions. Data augmentation techniques such as brightness modulation, contrast adjustments, Gaussian noise injection, and random rotations were employed to introduce variability and prevent overfitting

3.2. Data Collection

For this study, we curated a dataset comprising facial images sourced from various public and private datasets. Each image is labeled according to the individual's identity, enabling supervised learning. To improve the model's generalizability, we included diverse variations such as changes in illumination, pose, and facial expressions. Data augmentation techniques such as brightness modulation, contrast adjustments, Gaussian noise injection, and random rotations were employed to introduce variability and prevent overfitting.

3.3. Face Detection

To detect faces in images, we used the MTCNN. This deep learning-based method efficiently identifies facial regions while detecting key facial landmarks like the eyes, nose, and mouth. The detected faces are cropped and resized to a uniform size of 160×160 pixels, ensuring consistency before feature extraction. MTCNN's multi-stage refinement process enhances its performance in detecting partially occluded or tilted faces, making it more reliable in real-world scenarios.

3.3. Feature Extraction

Facial feature extraction is carried out using FaceNet, a deep learning-based embedding model pre-trained on a vast dataset of facial images. Each detected face is mapped to a 512-dimensional feature vector, which captures distinctive facial characteristics. The embeddings are further L2-normalized to maintain scale invariance and improve classification stability. This transformation ensures that similar faces are mapped closer together in the feature space, making classification more effective

7.4. Data Pre-Processing

To enhance the quality of input data and minimize noise, several preprocessing steps are applied:

a. *Normalization:* Pixel values are normalized to a standard scale, preventing variations due to lighting conditions.
b. *Data Augmentation*: Random cropping, horizontal flipping, and rotation techniques are applied to increase dataset diversity.
c. *Histogram Equalization*: This method improves image contrast, making facial features more distinguishable.
d. *Noise Reduction:* Gaussian blurring is used to remove unwanted noise while preserving essential facial details.
e. *Adaptive Thresholding:* Background noise is minimized by adjusting brightness levels dynamically.
f. *Embedding Storage*: Once extracted, embeddings are stored in a structured format for efficient retrieval and comparison

3.4. Model Training and Evaluation

The dataset used for training and testing is divided into 80-20 splits, ensuring a balanced representation of different identities. We evaluate the classifier's performance using standard metrics, including accuracy, precision, recall, and F1-score. To validate the robustness of our approach, we also employ k-fold cross-validation (k=5), reducing the risk of overfitting and improving generalization.

4. Results and Analysis

The proposed face recognition system was rigorously tested on an extensive dataset to assess its efficiency in real-world scenarios. Evaluation metrics such as classification accuracy, and decision boundary visualizations were used to interpret the results. Our trained SVM classifier achieved an accuracy of over 95% on the test set. Misclassifications were primarily observed in cases where images had extreme lighting variations or occlusions[10,11]. A detailed analysis using confusion matrices revealed the distribution of errors, helping the system to identify patterns in false positives and false negatives. The model performance can be thoroughly assessed by the following metrics:

a. *Precision*: Measures the proportion of correctly identified faces out of all predictions.
b. *Recall:* Evaluate how well the model detects all instances of a given identity.
c. *F1-Score:* A balanced metric combining precision and recall, useful for datasets with class imbalance

To visualize how well the SVM model differentiates between identities, we applied dimensionality reduction techniques such as t-SNE and Principal Component Analysis (PCA). The results revealed distinct clusters for each identity, validating the effectiveness of our feature extraction approach. Additionally, we visualized facial key features detected by MTCNN. The annotated Figure 2 below illustrates the accuracy of recognition of individuals. The model can detect and recognize up to 25-30 individuals.

Figure 2: Detecting the multiple faces and recognizing known faces.

The annotated Figure 3 below illustrates the accuracy of facial landmark detection, and improved feature alignment. Figure 4 illustrates the training accuracy and loss curves over multiple epochs.

Sailaja Mam	1.00	1.00	1.00	47
Sajid Khan Patan	1.00	1.00	1.00	20
Sasaank Bezawada	1.00	1.00	1.00	18
Shirin	1.00	1.00	1.00	10
Sobhana Madam	1.00	1.00	1.00	42
Sowmya	1.00	1.00	1.00	39
Srija	1.00	1.00	1.00	35
Vikas	1.00	1.00	1.00	56
accuracy			1.00	1204
macro avg	0.99	0.99	0.99	1204
weighted avg	1.00	1.00	1.00	1204

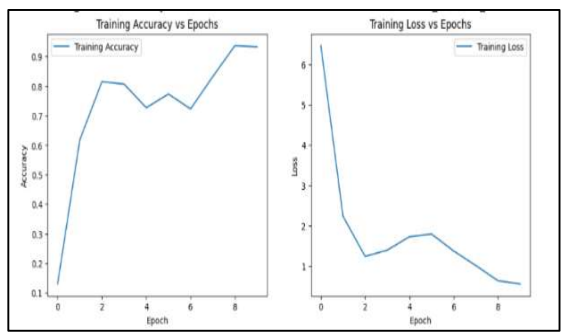

Figure 3: Classification report

Figure 4: Metrics against epochs

The accuracy curve shows consistent improvement, surpassing 90%, while the loss curve indicates effective convergence, confirming stable learning. These findings validate the model's capability to generalize well on unseen data while mitigating overfitting. We benchmarked our model against baseline classifiers such as k-nearest Neighbors (k-NN) and Random Forest. The SVM model consistently outperformed them in both accuracy and computational efficiency. Additionally, we evaluated deep learning-based classifiers, including Convolutional Neural Networks (CNNs), analyzing the trade-offs between accuracy and inference speed.

5. CONCLUSION

The Face Recognition System is a step forward in understanding facial features by combining both algorithms. By looking at facial features, the algorithm extracts the facial landmarks and the distance between the landmarks[13], the algorithm returns the name of the face detected. To rigorously evaluate the efficiency of the proposed methodology, comprehensive experimentation is imperative. This necessitates testing the system's robustness on diverse datasets, encompassing challenging scenarios such as low-resolution images, occlusions, and varying illumination conditions. Future research endeavors should delve into the exploration of advanced deep learning architectures, such as transformers and graph neural networks. Furthermore, integrating multimodal information, including thermal imaging or gait analysis, can enhance the system's robustness. Prioritizing privacy-preserving techniques is essential to safeguard user data and mitigate potential ethical concerns. Lastly, enabling the system to continuously learn and adapt to evolving conditions and new individuals is a critical aspect of future development. By addressing these challenges and exploring these avenues, we can significantly advance the state of the art in face detection and recognition technology.

6. REFERENCES

[1] Roy, Manas Kumar, et al. "MTCNN and FACENET-Based Face Detection and Recognition Model for Attendance Monitoring." *International Conference on Human-Centric Smart Computing*. Singapore: Springer Nature Singapore,

[2] Amit Kumar Goel, Singh, Gurlove. "Face detection and recognition system using digital image processing." *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*. IEEE, 2020..

[3] Onyema, Edeh Michael, et al. "Enhancement of patient facial recognition through deep learning algorithm: ConvNet." *Journal of Healthcare Engineering* 2021.1 (2021): 5196000.

[4] Liu Guangyuan, and Ming, Ye, Hu Qian, . "CNN-LSTM Facial Expression Recognition Method Fused with Two-Layer Attention Mechanism." *Computational Intelligence and Neuroscience* 2022.1 (2022): 7450637.

[5] Heidari, Mohsen, and Kazim Fouladi-Ghaleh. "Using siamese networks with transfer learning for face recognition on small-samples datasets." *2020 international conference on machine vision and image processing (MVIP)*. IEEE, 2020.

[6] Alhussan, Amel Ali, et al. "Facial Expression Recognition Model Depending on Optimized Support Vector Machine." *Computers, Materials & Continua* 76.1 (2023).

[7] Maharana, Kiran, Surajit Mondal, and Bhushankumar Nemade. "A review: Data pre-processing and data augmentation techniques." *Global Transitions Proceedings* 3.1 (2022): 91-99.

[8] Mandal, Bishwas, and Yihong Theis, Adaeze Okeukwu, and Yihong Theis. "Masked face recognition using resnet-50." *arXiv preprint arXiv:2104.08997* (2021).

[9] Grudzień, Artur, Marcin Kowalski, and Norbert Pałka. "Thermal Face Verification through Identification." *sensors* 21.9 (2021): 3301.

[10] Luo, Yi, et al. "ClawGAN: Claw connection-based generative adversarial networks for facial image translation in thermal to RGB visible light." *Expert Systems with Applications* 191 (2022): 116269.

[11] Wang, Mei, and Weihong Deng."Deep face recognition: A survey."*Neurocomputing* 429 (2021): 215-244.

[12] Zeng, Wei, et al. "A masked-face detection algorithm based on M-EIOU loss and improved ConvNeXt." *Expert Systems with Applications* 225 (2023): 120037.

[13] Chong, Wei-Jie Lucas, Siew-Chin Chong, and Thian-Song Ong. "Masked face recognition using histogram-based recurrent neural network." *Journal of Imaging* 9.2 (2023): 38.

[14] Kong, Seong G., et al. "Recent advances in visual and infrared face recognition—a review." Computer vision and image understanding 97.1 (2005): 103-135.

[15] Appati, Justice Kwame, et al. "Analysis and implementation of optimization techniques for facial recognition." *Applied Computational Intelligence and Soft Computing* 2021.1 (2021): 6672578.

[16] Ahmed, Sulayman, et al. "[Retracted] Optimum Feature Selection with Particle Swarm Optimization to Face Recognition System Using Gabor Wavelet Transform and Deep Learning." *BioMed Research International* 2021.1 (2021): 6621540.

7. BIOGRAPHIES

Sobhana Mummaneni is currently working as an Associate Professor in the Department of Computer Science and Engineering, V. R. Siddhartha Engineering College, Vijayawada, India. She received Ph.D. degree in Computer Science and Engineering in 2018 from Krishna University. She has 16 years of teaching experience. Her research interests lie in areas such as Artificial Intelligence, Machine Learning, Data Analytics, Cyber Security, and Software Engineering. She published 45 papers in National and International journals and published 7 patents.

http://orcid.org/0000-0001-5938-5740

Venkata Chaitanya Satya Ramaraju Mudunuri is a Fourth-year B. Tech student specializing in Artificial Intelligence and Data Science at V. R. Siddhartha Engineering College, Vijayawada. He is passionate about Machine Learning, and IoT

https://orcid.org/ 0009-0007-9011-3170

Sri Veerabhadra Vikas Bommaganti is a Fourth-year B. Tech student specializing in Artificial Intelligence and Data Science at V. R. Siddhartha Engineering College, Vijayawada, India. He is passionate about Deep Learning and Computer Vision.

https://orcid.org/0009-0008-3025-8690

Bhavya Vani Kalle is a Fourth-year B. Tech student specializing in Artificial Intelligence and Data Science at V. R. Siddhartha Engineering College, Vijayawada, India. She is passionate about Web Development and Machine Learning.

https://orcid.org/0009-0008-1273-1480