http://doi.org/10.35784/iapgos.xxxx *received: xx.xx.202x | revised: xx.xx.202x | accepted: xx.xx.202x | available online: 30.06.2024*

FACE RECOGNITION IN DENSE CROWD USING DEEP LEARNING APPROACHES WITH IP CAMERA

Sobhana Mummaneni[1], Venkata Chaitanya Satya Ramaraju Mudunuri[2], Sri Veerabhadra Vikas Bommaganti[2], Bhavya Vani Kalle[2], Novaline Jacob[3], Emmanuel Sanjay Raj[3]

[1]Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, [2] Velagapudi Ramakrishna Siddhartha Engineering College, Department of Artificial Intelligence and Data Science, [3] Advanced Data Processing Research Institute, Department of Space.

Abstract. *A facial recognition system is a biometric security and surveillance that has the potential to identify and monitor individuals in a crowded area. While monitoring a crowded environment manually is a tough and error-prone task. So, under such circumstances to improve security there is a need for a model that automatically detects and recognizes the faces of individuals. Automating facial recognition benefits in more efficient and accurate solution. This paper proposes a model "Face recognition in Dense crowd using Deep learning techniques". A sophisticated model that has the capability of detecting and recognizing faces in dense crowds was made possible by using deep learning techniques. In the case of live video as input data, the process includes dividing of video into frames and each frame is fed to the model. The algorithm MTCNN is used for face detection tasks, which accurately locates faces in the frames, and images and generates bounding boxes around the face as its output. Now the detected faces are fed as input to the model, if the faces in the frame are matched with the database, the name of the individual is displayed on the boundary box, else unknown will be displayed instead of the name of the person. FaceNet is used for face recognition tasks. The automated model of surveillance results in effective security and streamlined access control.*

Keywords: Face Detection, Multi-Task Cascaded Convolutional Neural Networks, Dense Crowd, FaceNet.

Introduction

Biometrics and surveillance are the basic ways of maintaining authentication and security which are made must in every workplace. 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 Deep Learning-based face recognition model that is trained to identify individuals regardless of occlusion. Facial recognition system maintains security, and surveillance regardless of crowd, occlusion [6].

The data collection phase is the beginning stage of the face recognition process which includes capturing a variety of faces of individuals, both normal and occluded faces. The dataset is created with 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 different circumstances and makes it work more efficiently than the traditional face recognition systems [15].

The proposed model utilizes DL techniques for instance Multi-Task Cascaded Convolutional Neural Networks (MTCNN) for the face detection process. Multi-Task Cascaded Convolutional Neural Networks (MTCNN) is primarily used for face detection tasks, it helps in accurately locating faces in images under complex conditions like occlusions, varied poses, and lighting changes. Face detection in Multi-Task Cascaded Convolutional Neural Networks (MTCNN) is done through a cascaded structure, this kind of approach makes the work model efficient. Its high frequency and speed make it suitable for surveillance and security in dense crowds.

The recognition phase includes techniques like FaceNet. The faces that are been detected in the previous stage are passed through FaceNet model. Depending on their recognition accuracy the system opts for the most reliable result, to boost the recognition process. This approach ensures the model can handle various faces and crowded conditions effectively.[17].

This entire process results in a video stream or an image with bounding boxes and facial landmarks. At the detection stage, the recognized individuals are identified by their names; otherwise, they are unknown. This provides identification and verification of an individual in a dense crowd.

1. Literature Review

Kong et al. [13] studied contemporary advancements in face recognition, focusing on 2D images in the visible and infrared spectra. While 2D visual recognition systems have matured, they face limitations in varying conditions. In contrast, infrared imaging offers robustness, especially for challenging scenarios. The paper presents deep learning models for recognition of faces in infrared environment, highlighting their accuracy and potential for security applications. Further research may refine convolutional architectures and evaluate performance with comparative metrics. Data accessibility is a current limitation in the field, but promising results suggest significant future improvements.

Edeh Michael Onyema et al. [21] conducted a study on improving facial expression recognition using a Convolutional Neural Network (ConvNet). They used training data from the FER2013 dataset, which includes seven facial expressions. Their approach significantly enhanced recognition accuracy without the computational overhead of deep-layered CNN models. Their computationally efficient model can be integrated with other systems, enhancing facial recognition accuracy, and future work may involve dataset expansion and incorporation of natural language processing for advanced automatic facial expression recognition in e-Health systems.

Ye Ming et al. [20] addresses limitations in existing facial expression recognition algorithms, focusing on the incorporation of attention mechanisms. A CNN-LSTM form the foundation for a novel CNN-ALSTM method, enhancing important region information mining. Additionally, a dual-layer attention mechanism is introduced. Comparative experiments on Fer2013 and processed CK+ datasets indicate 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. Future work involves training under extreme conditions, such as low light and occlusion.

Mohsen Heidari et al. [11] employed a siamese network approach with transfer learning from a VGG-16 model for face recognition in this paper. Two similar CNNs process pairs of face images, utilizing a similarity criterion to determine if they belong to the same person. The results demonstrate 95.62\% accuracy on the LFW dataset. To advance this research, exploring alternative CNN architectures within the siamese network, especially those capable of extracting both high and low-level features, is recommended. Furthermore, adopting the "triplet loss" method and employing data augmentation techniques may enhance performance, particularly for small-sample datasets.

Weijun Chen et al. [5] investigated face detection in object recognition, leveraging the rapid advances in deep learning. YOLO-face, built upon YOLOv3, uses specialized anchor boxes and a refined regression loss function to significantly boost accuracy while maintaining detection speed. Experiments on WIDER FACE and FDDB datasets highlight its superior performance over YOLO and its variants. YOLOv3 forms the base, adapted with specific anchor box scales, GIoU loss function, and network structure. The method achieves a balance between accuracy and speed, accommodating diverse scenarios and adjustments, like larger input images and scenario-specific anchor boxes. Face detection, a unique aspect of object recognition, benefits from this specialization, particularly in distinctive scenarios.

Shivalila Hangaragi et al. [10] conducted a study on face detection and recognition using Face Mesh. The model was trained on LWF dataset images and real-time captures. During testing, the model compared face landmarks with training images, labeling the person if matched or as "unknown" if not. Achieving 94.23\% accuracy, experiments included 1700 images from LWF and real-time data. Face Mesh was used for full face reconstruction and landmark detection, while accuracy was assessed with the BU3DFE dataset, considering various conditions and compared to existing methods.

Gurlove Singh et al. [24] explored face recognition, a critical element of individual authentication. The process involves two phases: face detection and recognition. The work examined the Eigenface and Fisherface 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.

Bishwas Mandal [18] et al. highlighted the urgent need for improved face recognition models to tackle the challenge of recognizing masked faces during the COVID-19 pandemic. Their research presents a sophisticated deep learning model, the ResNet-50 architecture, designed specifically for accurate masked face identification. While recognizing the constraints inherent in current convolutional networks when handling obscured faces, the study puts forth innovative strategies. These include cropping masked regions and delving into 3D learning to elevate performance, leveraging techniques such as data augmentation, domain adaptation, and the integration of a variety of CNN architectures and machine learning methods.

Artur Grudzien et al. [8] studied a novel approach to face verification in long-wavelength infrared radiation. The method combines two face images into a single double image and employs neural network-based classification. Testing was conducted on diverse thermal face databases. The method achieved an 83\% true acceptance rate, outperforming baseline methods by 20\%. The proposed technique, inspired by Siamese architecture, is adaptable to other spectral ranges and modalities. Results indicate the viability of longwave infrared face recognition, with potential for broader applications and architecture enhancements.

Yi Luo et al. [16] addressed the challenge of translating facial thermal images into RGB visible images. A novel approach, named ClawGAN, is proposed to improve image quality and feature preservation. It introduces the concept of mismatch metric (MM) and incorporates a claw-connected network structure to the generative adversarial network (GAN) framework. Comprehensive evaluations demonstrate that ClawGAN outperforms existing methods in terms of image quality and face recognition accuracy, particularly under varying illumination conditions.

Mei Wang et al. [26] explored the transformative role of deep learning in face recognition (FR) research. It traces the evolution of deep FR methods, network architectures, and loss functions, emphasizing the two categories of face processing approaches. The survey highlights the significance of different databases used in model training and evaluation. It extends to cover diverse deep FR applications, from cross-factor to industrial scenarios. Finally, it outlines the emerging technical challenges and promising directions in deep FR, marking its impact on the field.

Wei Zeng et al. [27] explored the evolution of feature extraction networks in classification algorithms, focused on convolutional neural networks (CNNs). While CNNs have typically used stacked residual blocks for downstream feature extraction, this study introduces the YOLOX detection framework, integrating aspects of the Swin-Transformer for more efficient classification. It addresses issues with IOU convergence and regression coordinate accuracy by introducing a novel penalty term, creating a robust algorithm for applications such as masked face detection in various_scenarios. Tested on standard datasets, the proposed model showcases significant improvements in average precision.

Shervin Minaee et al. [19] studied that face detection is a pivotal aspect of facial recognition and analysis systems. Earlier techniques used hand-crafted features in classifiers for face detection. However, Deep neural networks revolutionized image classification in 2012.. Deep learning-based frameworks have significantly enhanced face detection accuracy, making them more suitable for uncontrolled environments.

2. Methodology

This paper aims to provide an approach to detect faces from facial image. This section consists of two stages:

1. Detecting the faces from the image which identifies the positions of facial features.
2. Recognizing the detected face from the existing database (consists of faces of multiple persons).

2.1. Classical Methods for Face Detection

a) Haar Cascade Classifiers*:*

Haar Cascade is an algorithm for detecting object that efficiently identifies objects within images or video streams. It utilizes simple rectangular features called Haar features to capture intensity differences between adjacent image regions. These features are combined into a cascade structure, where each stage filters out irrelevant regions, leading to a rapid and accurate detection process[23]. By scanning images at different scales and locations, the algorithm can detect objects of varying sizes and positions, making it suitable for real-time applications like face detection, pedestrian detection, and more. It employs a cascade of simple classifiers, each trained to detect specific facial features. By utilizing Haar features, which capture intensity differences between adjacent image regions, the algorithm can efficiently identify edges, lines, and corners, essential for recognizing facial features. The integral image technique significantly speeds up the feature calculation process. The cascade structure ensures efficient processing by quickly rejecting non-face regions. Windows that successfully pass all stages are considered potential face detections.

b) HOG:

Histogram of Oriented Gradients is a approach for describing the feature used for detection of object. HOG is used for face detection by extracting features that are durable to variations in illumination and pose[25]. It divides the given image into small cells or parts and calculates the histogram of gradient orientations within each cell or part. These histograms capture the local shape and texture information of the face. By combining the histograms of neighbouring cells into larger blocks and applying normalization, HOG creates feature descriptors that are invariant to changes in lighting and geometric transformations. Machine learning classifiers, such as SVMs, can detect face pixels using descriptors.

2.2. Deep Learning based methods

a) MTCNN

MTCNN, a pretrained model specifically designed for face detection, excels in accurately identifying faces in diverse environments, including varying lighting conditions, poses, and occlusions [9]. It finds applications in facial detection, face tracking, and emotion recognition. The MTCNN is a convolutional neural network (CNN) comprising three layers of networks.

P_NET: P-Net stands for Proposal Network, which initially extracts the facial features to decide the bounding box. These three layers are employed to extract relevant features in a frame. These layers progressively learn to identify facial characteristics. The output from the final convolutional layer is then fed into a classifier, which determines whether the input image contains a human face.

R_NET: R-Net, stands for Refine Network, adds a 128 FCN (Fully Connected Network) after the last convolution layer in order to grasp more image features than the P-Net. It is stricter to select the features and deletes unnecessary candidate face area which doesn't affect good.

O_NET: O-Net stands for Output Network which is a complicated convolutional network with one more convolutional layer than the previous layer that is R-Net. It results in outputting the five final features from the face areas. Face detection and bounding box regression produce x, y, w, and h coordinates. x and y are the top-left corner coordinates, and w and h are the width and height, respectively. These values help identify the face boundaries in the image.

b) YOLO

Yolo, an object detection model, can detect various objects and classes. Unlike traditional methods that scan images multiple times, Yolo processes the entire image in a single pass, making it extremely fast. It divides the image into a grid and predicts bounding boxes and class probabilities for each grid cell. YOLO can detect multiple objects in a single image, making it suitable for applications like self-driving cars, surveillance systems, and robotics due to its high accuracy and real-time performance.

2.3. Bounding Box Regressions (Loss Functions)

Each stage in MTCNN performs bounding box regression to adjust the face coordinates. For bounding box prediction, MTCNN minimizes the MSE between the predicted and true bounding box coordinates:

$$L_{bbox} = \frac{1}{N}\sum_{i=1}^{N}((x_i - \widehat{x_i})^2 + (y_i - \widehat{y_i})^2 + (w_i - \widehat{w_i})^2 + \left(h_i - \widehat{h_i}\right)^2 \quad (1)$$

x_i, y_i, w_i, h_i: True bounding box coordinates (top-left corner coordinates and width or height).

$\widehat{x_i}, \widehat{y_i}, \widehat{w_i}, \overline{h_i}$: Predicted bounding box coordinates

N : Number of bounding boxes in a batch.

2.4. Face Classification (Binary Cross-Entropy Loss)

The networks use binary cross-entropy loss for face classification, distinguishing faces from non-faces:

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N}(a_i \log(\hat{a}_i) + (1 - a_i) \log(1 - \hat{a}_i) \quad (2)$$

a_i : Ground truth label (1 for face, 0 for non-face).

$\hat{a}_i$: Predicted probability of the face.

2.5 Total Loss

The final loss for each network is a combination of the classification and bounding box losses, as well as a landmark detection loss if landmarks are used:

$$L_{total}=\lambda_{cls} L_{cls}+ \lambda_{bbox} L_{bbox}+ \lambda_{landmark} L_{landmark} \quad (3)$$

where λ_{cls}, λ_{bbox}, $\lambda_{landmark}$ are weights that balance the different loss components. The proposed face detection model is mainly developed on MTCNN model, because the MTCNN significantly increases the accuracy of face detection and reduce false positives. Unlike classical methods [1], MTCNN effectively detects faces under various conditions, including non-frontal poses, complex lighting, and partial occlusions. The hierarchical architecture allows it to detect faces from multiple angles and in real-world environments. MTCNN provides facial landmarks for each detected face, which can be used for face alignment, recognition, and emotion analysis. This built-in feature simplifies the process and reduces the need for additional landmark detection algorithms. Along with various advantages, the major advantage is hardware compatibility. Because MTCNN requires a GPU for real-time applications, it can still run on CPUs with acceptable performance. Its accuracy and robustness make it worth the computational cost, especially in high-stakes applications like security, where reliable face detection is essential. This makes MTCNN one of the most versatile and reliable face detection algorithms for real-world applications.

2.6. Face Recognition

Face recognition involves comparing detected faces with existing ones. To resolve such issue, the need of more powerful and reliable face recognition model is high [2]. FaceNet, a deep learning model for face recognition, represents faces as compact, high-dimensional vectors (embeddings) instead of traditional methods like Eigenfaces or Fisherfaces. Face recognition algorithm uses Euclidean space to map face images to numerical embeddings for similarity comparison.

FaceNet model can be implemented on the faces of the size 160X160 [22]. So, the extracted faces are resized to a standard size (160x160). The model is trained to learn a function that maps face images to compact feature vectors, or embeddings. These embeddings capture the underlying facial features and can be used to compare faces. During inference, a new image is fed into the model, and its embedding is calculated. This embedding is then compared to the embeddings of known individuals in a database. By measuring the distance between embeddings, the model can determine if the new face matches any known identity. This approach has proven highly effective in face recognition, even under challenging conditions like varying lighting, poses, and occlusions [14].

FaceNet uses a triplet loss function to create embeddings, focusing on making embeddings of the same person close together in Euclidean space, while embeddings of different people are far apart.

2.7. Triplet Loss

In FaceNet, the triplet loss function is designed to separate positive pairs (same person) from negative pairs (different people) by a minimum margin α. Given a triplet of images (Anchor X, Positive P and Negative N), the triplet loss is defined by:

$$L_{triplet} = \sum_{i=1}^{N}[\| f(X_i) - f(X_i) \|_2^2 - \| f(X_i) - f(X_i) \|_2^2 + \alpha]_+ \quad (4)$$

f(x): Embedding function mapping an input image x to a 512-dimensional embedding vector.

|| . ||2: Euclidean (L2) distance.

α: Margin that separates positive and negative pairs.

[·] +: Ensures the loss is zero when the difference is less than α meaning the positive pair is sufficiently close and the negative pair is sufficiently far.

This loss encourages the distance between a positive and an anchor to be smaller than the distance between a negative and an anchor by at least α. Here positive refers same person and negative refers different person.

2.8. Embedding Similarity Measurement

For recognition, FaceNet calculates the Euclidean distance between embeddings to determine similarity:

$$d\,(A, B) = ||\, f(A) - f(B)\, ||2 \qquad (5)$$

If *d (A, B) < threshold*, then A and B are considered to be the same person.

2.9. Comparison Metric:

Sometimes, cosine similarity is helpful to measure the similarity between embeddings instead of Euclidean distance:

$$cosine\_similarity(A, B) = \frac{f(A).f(B)}{\|f(A)\|\,\|f(B)\|} \qquad (6)$$

2.10. Data Collection and Processing

Data Preparation: The dataset used in this project consists of labeled images of various individuals, organized in subdirectories by person or identity. For the efficient performance the dataset has to be more accurate and maintains the standards like image quality, light conditions (bright and dim), different angles (+45 degrees to -45 degrees). Each subdirectory represents a unique individual or class label, containing multiple images with varying conditions such as lighting, facial expressions, and angles. This structure allows the model to learn diverse features for each individual, aiding in robust recognition.

In this dataset, there are 26 classes in which each class varies from 100 to 150 images. It involves various angles like frontal and side faces. Along with that there are various light conditions. The figure 1 shows front, left and right side of the faces. Each image is loaded using OpenCV, which by default reads images in BGR format. Since most deep learning models are trained on RGB images, a conversion step from BGR to RGB is essential. This step aligns the color channels with the format used by the FaceNet model, improving embedding accuracy. Before extracting faces, the images are resized and preprocessed to ensure uniformity across samples. Each face is detected using the MTCNN (Multi-task Cascaded Convolutional Networks) detector, which is proficient in locating faces and distinguishing features.

The MTCNN operates in three stages:

- It first identifies face candidates through a simple, fast network.
- A more refined network further filters the candidates.
- A final network adjusts the bounding box coordinates to capture the face accurately.

2.10. Face cropping and Resizing

Fig . 1. Dataset Images

After detecting the face, the bounding box is like a magic frame that cuts out the face region from the image[12]. This region is resized to a target dimension of 160x160 pixels, ensuring consistent input size for the FaceNet model. This standard size was chosen because it balances computational efficiency and the retention of facial details needed for recognition. In Fig 1, the dataset also contains some partial occlusions.

2.11. Face Embedding Generation

Embedding Process: For each detected face, the FaceNet model generates a 512-dimensional embedding, which is a numerical representation of the facial features. Each dimension represents a specific feature extracted by the model, such as eye shape, nose length, or jaw structure. The high dimensionality allows the model to capture subtle differences between faces, enhancing recognition accuracy. Each face is normalized to ensure consistent input, which includes converting pixel values to floating-point numbers and scaling between [0, 1].

This process reduces variations from lighting and contrast. Each dimension represents a specific feature extracted by the model, such as eye shape, nose length, or jaw structure. The high dimensionality allows the model to capture subtle differences between faces, enhancing recognition accuracy.

The embedding vector simplifies the comparison process during recognition. Instead of comparing pixel-by-pixel, which is inefficient and error-prone, embeddings allow the use of efficient distance metrics (e.g., Euclidean distance) to determine if two faces belong to the same person[3].

Storage and Retrieval: These embeddings are stored in a serialized format for efficient retrieval. This file acts as a face “database,” allowing the model to compare incoming face embeddings with stored embeddings during the recognition phase. Since the storing the face is not safe, it saves the embeddings data.

To save computational time, embeddings are stored in the pickle format. This allows the model to quickly retrieve and match faces without recalculating embeddings each time.

2.12. Data Augmentation Techniques

To make the model more reliable, more images are added by messing with them, like changing the lighting, flipping them, and changing how they look, such as:

Horizontal Flipping: This mirrors the face horizontally, aiding in recognition under different orientations.

Brightness Adjustment: Varying the brightness prepares the model for images taken under different lighting conditions.

Slight Rotation: Small rotations prepare the model for slight tilts and angles in real-world images.

2.13. Model Training

For the model training on dataset, model use the SVC classifier from SVM module. The train split of size 75% and test split of size 25%. The model use the shuffle technique since if it don’t use the shuffle technique the order of samples in the dataset will directly affect how they are split into the training and testing sets.

For example, if the dataset is sorted based on some criteria (e.g., labels), the training and testing sets will contain consecutive samples belonging to the same category. The kernel we used for the svm model is linear. This process entails identifying the optimal hyperplane that effectively separates the various classes represented by the training data points within the high-dimensional feature space, which is defined by the embedded facial vectors. SVM classifier [4] trains on training data after splitting data into training and testing sets. The trained model can subsequently be employed to generate predictions for novel facial embeddings**.**

3. Results

The results of our face detection and recognition project demonstrate the effectiveness of combining MTCNN for detection with FaceNet and an SVM for recognition. During the detection phase, MTCNN achieved high accuracy, reliably identifying and isolating faces in a variety of conditions, including variations in lighting, angles, and facial expressions. While MTCNN's performance was slightly slower than some other face detection models, it outperformed many in terms of accuracy, proving particularly useful for applications prioritizing detection quality over speed [7].

In the recognition phase, the SVM classifier trained on FaceNet-generated embeddings demonstrated impressive results. The model achieved high accuracy on the training set, reflecting its capability to learn and distinguish between different individuals. On the testing set, the accuracy remained high, with only a minor reduction compared to the training accuracy, indicating the model's ability to generalize well.

Precision, recall, and F1-scores across different classes were strong, reflecting the model's balanced performance in correctly identifying individuals and avoiding misclassifications. Additionally, a confusion matrix provided insight into any occasional misclassifications, although most individuals were accurately identified, as indicated by high diagonal values.

The results also highlight several advantages and limitations of this system. The embedding-based approach is not only accurate but also scalable, making it well-suited for real-world applications that require quick and reliable recognition across large databases. Moreover, using embeddings enhances privacy, as these numerical representations of faces are not easily reversible to reconstruct the original image, thereby providing a degree of security in data storage.

In Fig 2, the MTCNN and FaceNet combination achieved excellent results, with the SVM classifier maintaining high accuracy and efficiency in recognizing faces. The model's performance metrics—such as precision, recall, and F1-score—alongside its successful generalization to testing data, affirm its reliability for practical applications. Overall, the results suggest that this face recognition system is a viable solution for real-world use, particularly in applications requiring high accuracy and secure data handling, though further fine-tuning and expanded datasets could enhance its robustness in more diverse scenarios.

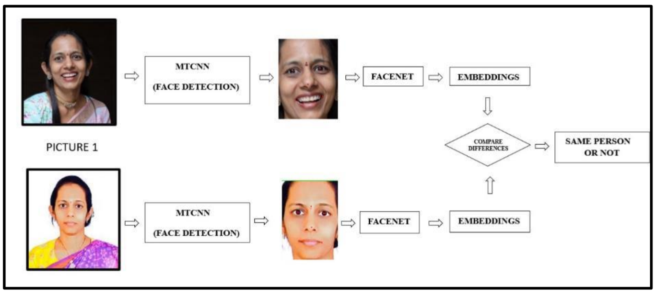

Fig 2: Process flow of Face Detection and Recognition

In Fig 3, the model can be able to detect and recognizes the faces provided. If a face is matched with the database clusters then the bounding box is characterized with green color having the name and accuracy as attributes, else the bounding box is characterized with blue color having "UNKNOWN" label and accuracy as attributes.

Precision, Recall, and F1-Score: Each individual in the dataset was evaluated with precision, recall, and F1-score. The model achieved high accuracy, with performance metrics of 1.00 for most individuals. This indicates minimal false positives and negatives, effectively recognizing faces. Most individuals have a precision of 1.00, which shows that the classifier made very few incorrect identifications.

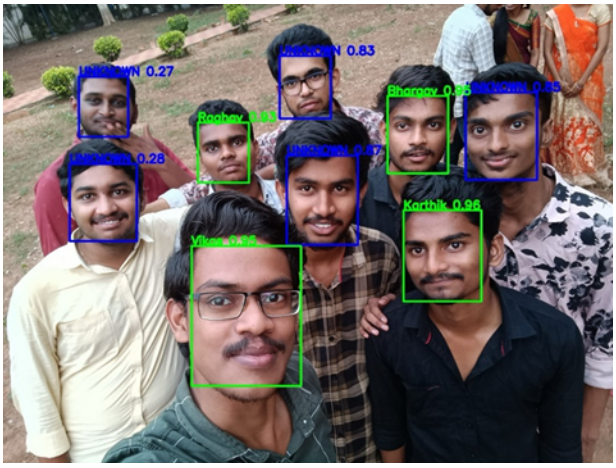

Fig 3: Face Recognition in Dense Crowd

$$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP)+False\ Positives(FP)} \tag{7}$$

$$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP)+False\ Negatives(FN)} \tag{8}$$

Recall measures how well the model can find all the faces in the dataset. If the recall is high for each person, it means the model did a great job of recognizing almost all the faces without missing anyone.

F1-Score is like a balance between precision and recall. The high F1-scores for each person mean the model was super accurate and reliable.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

The confusion matrix visually represents a face recognition model's performance on a test dataset. Each row shows the actual identity, while each column shows the predicted identity. The diagonal shows correct predictions, while off-diagonal elements indicate misclassifications.

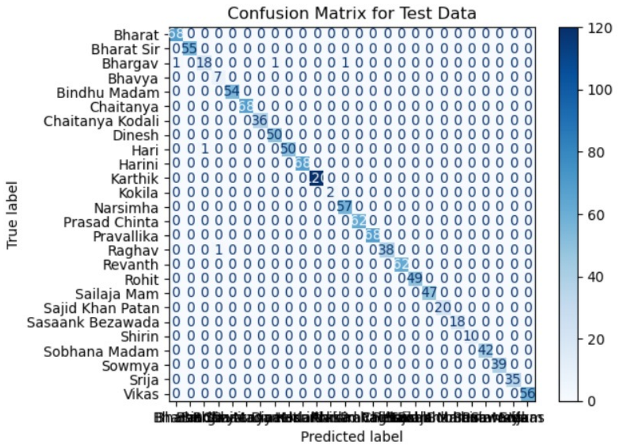

Fig 4 : Confusion Matrix for Face Recognition Model on Test Dataset

In Fig 4, the color gradient from light to dark blue indicates the frequency of predictions, with darker shades representing higher counts. Here, we see a high concentration along the diagonal, indicating the model's strong performance, as most faces are correctly classified. Misclassifications are minimal, reflecting high accuracy and reliability in distinguishing between individuals in the dataset.

4. Future Work

To rigorously evaluate the efficacy 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. Additionally, assessing the system's real-time performance and its resilience against adversarial attacks is crucial.

Future research endeavours 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. Prioritising 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.

References

[1] Ali, Waqar, et al. "Classical and modern face recognition approaches: a complete review." Multimedia tools and applications 80 (2021): 4825-4880 [http://dx.doi.org/10.1007/s11042-020-09850-1].

[2] Appati, Justice Kwame, et al. "Analysis and implementation of optimization techniques for facial recognition." Applied Computational Intelligence and Soft Computing 2021.1 (2021): 6672578 [http://dx.doi.org/10.1155/2021/6672578].

[3] Archana, M. C. P., C. K. Nitish, and Sandhya Harikumar. "Real time face detection and optimal face mapping for online classes." Journal of Physics: Conference Series. Vol. 2161. No. 1. IOP Publishing, 2022 [http://dx.doi.org/10.1088/1742-6596/2161/1/012063].

[4] Chandra, M. A., & Bedi, S. S. (2021). Survey on SVM and their application in image classification. International Journal of Information Technology, 13(5), 1-11 [https://doi.org/10.1007/s41870-017-0080-1].

[5] Chen, Weijun, et al. "YOLO-face: a real-time face detector." The Visual Computer 37 (2021): 805-813 [http://dx.doi.org/10.1007/s00371-020-01831-7]

[6] 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 [http://dx.doi.org/10.3390/jimaging9020038].

[7] Coe, James, and Mustafa Atay. "Evaluating impact of race in facial recognition across machine learning and deep learning algorithms." Computers 10.9 (2021): 113 [https://doi.org/10.3390/computers10090113]

[8] Grudzień, Artur, Marcin Kowalski, and Norbert Pałka. "Thermal Face Verification through Identification." sensors 21.9 (2021): 3301 [https://doi.org/10.3390/s21093301].

[9] Gu, M., Liu, X., & Feng, J. (2022). Classroom face detection algorithm based on improved MTCNN. Signal, Image and Video Processing, 16(5), 1355-1362. [https://doi.org/10.1007/s11760-021-02087-x]

[10] Hangaragi, Shivalila, Tripty Singh, and N. Neelima. "Face detection and Recognition using Face Mesh and deep neural network." Procedia Computer Science 218 (2023): 741-749 [http://dx.doi.org/10.1016/j.procs.2023.01.054].

[11] 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 [http://dx.doi.org/10.1109/MVIP49855.2020.9116915].

[12] Khan, Amjad Rehman, et al. "Face detection in close-up shot video events using video mining." Journal of Advances in Information Technology 14.2 (2023) [http://dx.doi.org/10.12720/jait.14.2.160-167].

[13] 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 [http://dx.doi.org/10.1016/j.cviu.2004.04.001].

[14] Kumar, Ashu, Munish Kumar, and Amandeep Kaur. "Face detection in still images under occlusion and non-uniform illumination." Multimedia Tools and Applications 80 (2021): 14565 14590. [https://doi.org/10.1007/s11042-020-10457-9]

[15] Lu, Peng, Baoye Song, and Lin Xu. "Human face recognition based on convolutional neural network and augmented dataset." Systems Science & Control Engineering 9.sup2 (2021): 29-37 [https://doi.org/10.1080/21642583.2020.1836526].

[16] 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 [http://dx.doi.org/10.1016/j.eswa.2021.116269].

[17] Mamieva, Dilnoza, et al. "Improved face detection method via learning small faces on hard images based on a deep learning approach." Sensors 23.1 (2023): 502 [https://doi.org/10.3390/s23010502]

[18] Mandal, Bishwas, Adaeze Okeukwu, and Yihong Theis. "Masked face recognition using resnet-50." arXiv preprint arXiv:2104.08997 (2021) [https://doi.org/10.48550/arXiv.2104.08997].

[19] Minaee, Shervin, et al. "Going deeper into face detection: A survey." arXiv preprint arXiv:2103.14983 (2021) [https://doi.org/10.48550/arXiv.2103.14983]

[20] Ming, Ye, Hu Qian, and Liu Guangyuan. "CNN-LSTM Facial Expression Recognition Method Fused with Two-Layer Attention Mechanism." Computational Intelligence and Neuroscience 2022.1 (2022): 7450637 [http://dx.doi.org/10.1155/2022/7450637].

[21] Onyema, Edeh Michael, et al. "Enhancement of patient facial recognition through deep learning algorithm: ConvNet." Journal of Healthcare Engineering 2021.1 (2021): 5196000 [http://dx.doi.org/10.1155/2021/5196000].

[22] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823). [https://doi.org/10.1109/CVPR.2015.7298682]

[23] Shetty, Anirudha B., and Jeevan Rebeiro. "Facial recognition using Haar cascade and LBP classifiers." Global Transitions Proceedings 2.2 (2021): 330-335 [https://doi.org/10.1016/j.gltp.2021.08.044].

[24] Singh, Gurlove, and Amit Kumar Goel. "Face detection and recognition system using digital image processing." 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). IEEE, 2020 [http://dx.doi.org/10.1109/ICIMIA48430.2020.9074838].

[25] Surasak, T., Takahiro, I., Cheng, C. H., Wang, C. E., & Sheng, P. Y. (2018, May). Histogram of oriented gradients for human detection in video. In 2018 5th International conference on business and industrial research (ICBIR) (pp. 172-176). IEEE [https://doi.org/10.1109/ICBIR.2018.8391187].

[26] Wang, Mei, and Weihong Deng. "Deep face recognition: A survey." Neurocomputing 429 (2021): 215-244 [https://doi.org/10.1109/SIBGRAPI.2018.00067].

[27] Zeng, Wei, et al. "A masked-face detection algorithm based on M EIOU loss and improved ConvNeXt." Expert Systems with Applications 225 (2023): 120037 [http://dx.doi.org/10.1016/j.eswa.2023.120037].

Prof. Sobhana Mummaneni
e-mail: sobhana@vrsiddhartha.ac.in

Sobhana Mummaneni 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

Eng. Mudunuri Venkata Chaitanya Satya Ramaraju
e-mail: satyaramaraju1234@gmail.com

Mudunuri V CH S Raju 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 ML and IoT
https://orcid.org/ 0009-0007-9011-3170

Eng. Bommaganti Sri Veerabhadra Vikas
e-mail: vikas2004bommaganti@gmail.com

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

Eng. Kalle Bhavya Vani
e-mail: bhavyavanikalle@gmail.com

Bhavy 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

Novaline Jacob
e-mail: novalinejacob@gmail.com

Novaline Jacob, MSc., MTech., Ph.D, is a Scientist at Advanced Data Processing Research Institute (ADRIN), Dept. of Space, Secunderabad. She has acquired her master's degree in Applied Geology (MSc. Applied Geology) and Remote Sensing (MTech. Remote Sensing) from Anna University, Chennai.
https://orcid.org/0009-0007-8627-9231

Emmanuel Sanjay Raj
e-mail: esanjayraj@gmail.com

K.C. Emmanuel Sanjay Raj is heading Video Analytics Section at ADRIN, Dept. of Space. He received degree in MSc(physics) from Osmania University,1994 and MS (Computer Science),2002. His research interests include Computer Vision, Intelligent video surveillance systems, Content-based Image retrieval, Digital Watermarking, Visual Cryptography and Steganalysis.
https://orcid.org/0009-0003-6389-6625