Facial Recognition Using YOLO object detection and FaceNet for features extraction

Aryan Tyagi, Aviral Rastogi, Anshika, Harsh Gupta, Diksha Arya

aryan.tyagi.cs.2019@miet.ac.in, aviral.rastogi.cs.2019@miet.ac.in, anshika.om.cs.2019@miet.ac.in, harsh.gupta.cs.2019@miet.ac.in, diksha.arya@miet.ac.in

Department of Computer Science and Engineering
Meerut Institute of Engineering and Technology, Meerut

Abstract – Facial verification is a critical application in various industries such as security, law enforcement, and access control systems. In this project, we propose a face verification system that combines state-of-the-art technologies, including YOLO (You Only Look Once) for object detection, FaceNet for facial feature extraction, and ArcFace for model training. The system utilizes YOLO to locate and extract faces from input images, and then FaceNet to extract high-quality facial features. We then use ArcFace to train our model for improved accuracy and robustness. The system calculates the similarity between faces using Euclidean distance, a widely used metric for measuring distance in high-dimensional spaces. The proposed methodology involves collecting a large dataset of images containing faces, preprocessing the data, training our model, and evaluating its performance using various metrics such as accuracy, precision, and recall. By following this methodology, we can build a robust and accurate face verification system that can be deployed in real-world scenarios to improve efficiency and effectiveness.

Keywords – Machine Learning, Deep Learning, Data Training, ArcFace, CNN, You Only Look Once (YOLO), FaceNet, and Euclidean distance.

1. Introduction

Face verification is the process of determining whether or not two facial images belong to a single individual. It is a necessary component of many applications, such as security systems, access control systems, and digital identity verification.
Deep learning-based techniques have demonstrated outstanding performance in face verification tasks in recent years. In this research, we propose a face verification system that integrates YOLO object detection, FaceNet face recognition, and the Euclidean distance metric to achieve high accuracy in face verification [1][2].

The suggested system begins by detecting faces in input photos using the YOLO object detection algorithm. YOLO is well-known for its ability to detect objects quickly and precisely. The recognized faces are then trimmed and aligned to reduce pose and illumination differences. Following that, the aligned faces are fed into a pre-trained FaceNet model, which maps each face image to a high-dimensional feature vector [3][4].

To compare the similarity of two feature vectors, we use the Euclidean distance metric, which computes the straight-line distance between two points in a high-dimensional space. If the distance between two feature vectors is less than a specific threshold, we consider them to be a match, indicating that the two face photos are from the same individual. If the distance is greater than the threshold, the two face photos are classified as non-matches, suggesting that they belong to separate people [5][6].

ArcFace loss function has been used to improve the accuracy of the face recognition model. The ArcFace loss function is designed to maximize the angular margin between the features of different classes [7][8].

2. Literature Review

Numerous studies have been conducted, and some of them have inspired us to do this study. Below are brief summaries of a few of those recent, important studies:

Insaf Adjabi et al, [1], have presented a 2D and 3D approach to study constrained and unconstrained face recognition. 2D approach is accomplished in controlled environment, where factors like lighting, point of view, and the distance between the camera and subject are all under our control. A certain level of maturity and extremely high rated of recognition were found for 2D approaches. 3D approach was suggested as an alternative solution, Due to the fact that 3D data is independent of position and illumination, recognition algorithms are more effective when using it. The paper discusses the history of facial recognition technology, the newest cutting-edge methods, and potential future advancements.

Yassin Kortli et al, [2], have proposed various techniques including local, holistic, and hybrid approaches, which describe a face image using all of the facial characteristics or just a few of the face image features. The benefits and drawbacks of each technique's schemes are listed in the paper along with their robustness, accuracy, complexity, and discrimination in order to provide a thorough comparison between them. The most popular datasets, including those for supervised and unsupervised learning, are outlined. Along with the context of the experiments and challenges that these techniques were used to address, the numerical findings of the most intriguing techniques are presented.

Murat Taskiran et al, [3], have presented the features of biometric system. The procedures used in the literature to collect and categorize facial biometric data have been compiled in this paper. They provide a classification of image-based and video-based face recognition techniques, a rundown of significant historical advancements, and an explanation of the key processing steps. A review of well-known data sets that researchers have employed for face identification is also provided. They also discuss the most recent facial recognition techniques based on deep learning and suggest future research avenues.

Ashu Kumar et al, [4], have presented a detailed examination of the many techniques looked into for face detection in digital photos. This essay also explores several face detection challenges and applications. The properties of numerous common face detection databases are also supplied. In order to create a reliable face detection system, they arrange special discussions on the practical aspects. They conclude this paper with several promising lines of inquiry.

Iacopo Masi et al, [5], have presented the major developments in deep face recognition and, more generally, in learning face representations for verification and identification are summarized. The survey presents the principle in a way that is organized and clear, state-of-the-art (SOTA) face recognition techniques appearing in top computer vision settings within the last five years. The survey is divided into several sections that stick to a typical facial recognition pipeline: (A) Which open data sets have been used in SOTA system training, (B) the facial preprocessing component (detection, alignment, etc.), (C) the transfer learning architecture and loss functions, (D) face detection technology for identification and verification.

Anagha S. Dhavalikar et al, [6], have proposed an Automatic Facial Expression Recognition System (AFERS). The suggested method consists of three steps: (A) face detection, (B) feature extraction, (C) facial expression recognition. The first phase of face detection entails detecting skin color using the YCbCr color model, compensating for lighting to achieve uniformity on the face, and performing morphological operations to retain the necessary face portion. The first phase's output is used to extract facial characteristics such as the eyes, nose, and mouth using the AAM (Active Appearance Model) technique. The simple Euclidean Distance technique is used in the third step, automatic facial expression recognition. This technique compares the Euclidean distance between the feature points of the training and query images. The output image expression is determined by the minimal Euclidean distance.

Madan Lal et al, [7], have presented different methods that can be used to build recognition systems. However, the procedures that use the iris and fingertips the most frequently. To access the system, these call for user interaction or involvement. Additionally, modern methods offer member access without its involvement. Face recognition databases are different for controllable images compared to uncontrollable videos, using LFT for photos and YTM for films. Three major parts make a face recognition system: (A) pre-processing, (D) feature selection, and (C) Classification.

Song Zhou et al, [9], have proposed the process flow for 3D face recognition consists of two phases and five steps. Preprocessing is used to "clean up" the 3D data that are obtained during the training phase. The characteristics that could be utilized to distinguish between distinct faces are then found by processing the data using feature extraction methods. Each face's features are then added to the feature database. The investment, preprocessing, and extraction of features phases for the target face in the testing phase are the same as those with in training phase. The target face's features are compared to those of faces that have been saved in the feature database to determine match scores during the feature matching step. We would assert that the target face is recognized when the match score is high enough.

By reviewing the previously mentioned works, we become aware of the significance of Face recognition System and its applications.

3. Proposed Methodology

First step for face verification is to Collect images of faces for both positive and negative pairs. Positive pairs are made up of photographs of the same person, whereas negative pairs are made up of images of

different people.
Then Preprocess the collected face images by cropping and aligning them to remove any variations in pose and illumination. This step ensures that the face images are normalized and suitable for feature extraction [9][10].
Now, Use YOLO object detection to detect faces in the input images. YOLO can quickly and accurately detect objects in images, including faces.
Extract the features of the detected faces using a pre-trained FaceNet model. FaceNet maps each face image to a high-dimensional feature vector that represents the unique characteristics of the face [11][12].
After extracting facial features calculate the similarity between the feature vectors of two face images using Euclidean distance metric. If the distance between the two feature vectors is below a certain threshold, we classify them as a match, indicating that the two face images belong to the same person. If the distance is above the threshold, we classify them as non-matches, indicating that the two face images belong to different people [13][14].
Below diagram depicts the Euclidean distance:

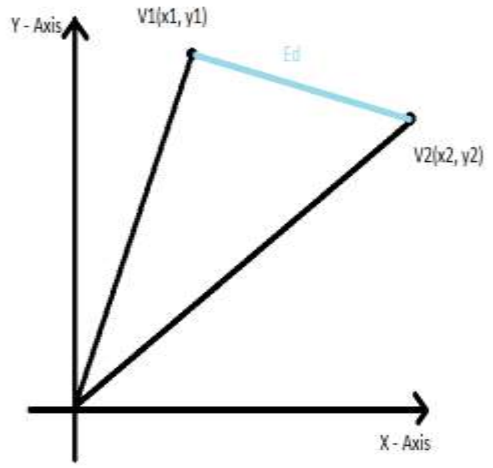

Fig. 1 – Euclidean Distance

Below is the formula for calculation:

$$\sqrt{\sum_{i=1}^{z} (V1i - V2i)^2}$$

Train a deep neural network using the ArcFace loss function to learn more discriminative embeddings for face recognition. By training the network with ArcFace loss, the embeddings learned by the network are more discriminative and better suited for face recognition tasks. Since the most important factor for building such system is using an appropriate loss function. ArcFace or Additive Angular Margin Loss is being used for discriminating two facial images due to inability of traditional softmax loss function for large intra facial variations, below is the formula for ArcFace loss function calculation [15][16][17].

$$L_3 = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)}}{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos\theta_j}}$$

At last, evaluate the performance of the face verification system using standard metrics, including accuracy, precision, recall, and F1 score [18][19].

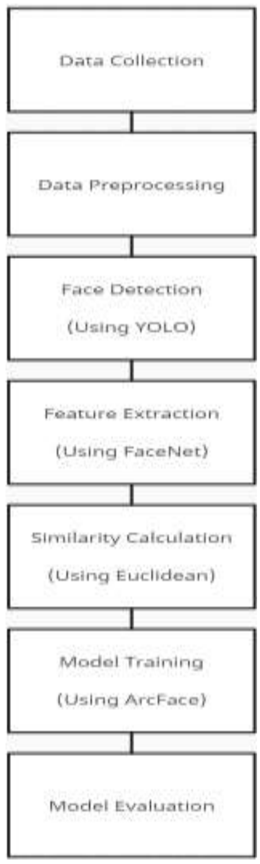

Fig. 2- Data Flow Diagram

After analyzing the results basis, the above factors, results of the services are returned[20].
Below are the services in this system:

- Face recognition service (Face identification)
- Face detection service
- Face verification service

Face detection service is used to detect all faces in the image. It doesn't recognize faces, just finds them on the image [21].

A face recognition service is one that uses facial recognition. This implies that you must first add recognized faces to the faces collection before identifying unfamiliar faces among them. The service returns the faces that are most similar to any unknown face you input. Additionally, the face recognition service has a validating endpoint so that you can confirm that the person you collected faces for is the right one [22].

Face verification service is a service that is used to confirm that this individual is the right one. The service compares two faces you send to the rest endpoint and returns their similarity [23][24]

4. Result and Discussions

This section describes the result analysis of the proposed work. This paper presented results of classification process in terms of accuracy, precision, recall and F1-Score. Every classification method requires on good accuracy for accurate classification, which is dependent on how well image characteristics are retrieved [25][26].

Accuracy: The percentage of values that were successfully categorized is calculated using accuracy. It indicates the frequency of correct classifications made by our classifier. It is the product of all true values divided by all values.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Precision governs how well the model classifies positive values. By dividing the total projected positive values by the actual positive values, it is calculated [27].

$$Precision = \frac{TP}{TP + FP}$$

Recall: It is employed to assess how well the model is able to estimate positive results. How often does the model predict positive values correctly? The calculation is done by dividing the total number of positive values by the positive values themselves [28].

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: It is the harmonic mean of Recall and Precision. When you need to consider both Precision and Recall, it is helpful.

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Fig. 3 – Sample images

We have performed our proposed work and calculated the confusion matrix which is shown in table 1.

Input Images (N) = 10	Sample Person 1	Sample Person 2	Sample Person 3	Sample Person 4	Sample Person 5
Sample Person 1	9	0	0	0	1
Sample Person 2	0	10	0	0	0
Sample Person 3	0	0	10	0	0
Sample Person 4	1	1	0	8	0
Sample Person 5	0	0	0	0	10

Table 1 - Demonstrates the confusion matrix for our proposed work.

Based on this above computed confusion matrix the performance metrics are calculated such as Accuracy, precision, recall and F1 Score.

Performance Metrics (%)	Person 1	Person 2	Person 3	Person 4	Person 5
Accuracy	90.9	100	100	80	100
Precision	90	90.9	100	100	90.9
Recall	90	100	100	80	100
F1-Score	90	95.23	100	88.88	95.23

Table 2 - Demonstrates the results of performance metrics calculated from proposed classification technique in %.

5. Conclusion

In conclusion, we have proposed a face verification system that combines state-of-the-art technologies to accurately and efficiently verify the identity of individuals using facial recognition. Our system utilizes YOLO for object detection, FaceNet for facial feature extraction, and ArcFace for model training. By using these cutting-edge technologies, we can detect and extract faces from input images, extract high-quality facial features, and train our model for improved accuracy and robustness. The system uses Euclidean distance to measure the similarity between faces and can be deployed in real-world scenarios such as access control systems, security applications, and law enforcement. The proposed methodology involves collecting a large dataset of images containing faces, preprocessing the data, training our model, and evaluating its performance using various metrics. By following this methodology, we can build a reliable and accurate face verification system that can improve the overall efficiency and effectiveness of various applications. In the future, we can further improve our system by incorporating additional technologies and techniques to enhance its performance and reliability.

6. References

[1] Insaf Adjabi, Abdeldjalil Ouahabi, Amir Benzaoui, Abdelmalik Taleb-Ahmed, “Past, Present, and Future of Face Recognition: A Review”, Journal Electronics 2020, *9*(8), 1188. Published: 23 July 2020.

[2] Yassin Kortli, Maher Jridi, Ayman AlFalou, Mohamed Atri, “Face Recognition Systems: A Survey”, Journal, Sensors 2020, 20(2), 342. Published: 7 January 2020.

[3] Murat Taskiran, Nihan Kahraman, Cigdem Eroglu Erdem, “Face recognition: Past, present, and future (a review)”, Digital signal processing Journal, Published: 16 July 2020.

[4] Ashu Kumar, Amandeep Kaur and Munish Kumar, “Face detection techniques: a review”, Artif Intell Rev 52, 927-928(2019). DOI: 10.1007/s10462-018-9650-2.

[5] Lacopo Masi, Yue Wu, Tal Hassner, Prem Natarajan, “Deep Face Recognition: A Survey”, 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images, Parana, Brazil, Doi: 10.1109/SIBGRAPI.2018.00067.

[6] Anagha S. Dhavalikar, R.K. Kulkarni, “Face detection and facial expression recognition system”, 2014, International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, Doi: 10.1109/ECS.2014.6892834.

[7] Madan Lal, Kamlesh Kumar, Rafaqat Hussain, “Study of Face Recognition Techniques: A Survey”, June 2018, International Journal Advanced Computer Science and Applications 9(6), DOI: 10.14569/1JACSA.2018.090606.

[8] Muhtahir Oloyede, Gerhard P. Hancke, Herman Myburgh, “Improving Face Recognition Systems Using a New Image Enhancement Technique, Hybrid Features and the Convolutional Neural Network”, November 2018, IEEE, DOI: 10.1109/ACCESS.2018.2883748.

[9] Song Zhou, Sheng Xiao, “3D face recognition: a survey”, Hum. Cent. Comput. Inf. Sci. 8, 35(2018), Doi: 10.1186/s13673-018-0157-2.

[10] Waqar Ali, Wenhong Tian, Salah Ud Din, Desire Iradukunda, Abdullah Aman Khan, “Classical and Morden Face Recognition Approaches”, Multimed Tools Appl 80, 4825-2880 (2021), DOI: 10.1007/s11042-020-0985-1.

[11] Mei Wang, Weihong Deng, “Deep Face Recognition: A Survey”, January 2021, DOI: 10/1016/j.neucom.2020.10.081.

[12] Shaveta Dargan, Munish Kumar, Maruthi Rohit Ayyagari, Gulshan Kumar, “A Survey of Deep Learning and Its Applications”, Arch computat Methods Eng 27, DOI: 10.1007/s11831-019-09344.

[13] Prentza, S. Maglavera, “Delivery of healthcare services over mobile phones”. WHO Guidelines for medicine donations

[14] Mohseni, S., Yang, F., Pentyala, S., Du, M., Liu, Y., Lupfer, N., ... & Ragan, E. (2021, May). Machine learning explanations to prevent overtrust in fake news detection. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 15, pp. 421-431).

[15] Narayan, Vipul, et al. "Enhance-Net: An Approach to Boost the Performance of Deep Learning Model Based on Real-Time Medical Images." Journal of Sensors 2023 (2023).

[16] Babu, S. Z., et al. "Abridgement of Business Data Drilling with the Natural Selection and Recasting Breakthrough: Drill Data With GA." Authors Profile Tarun Danti Dey is doing Bachelor in LAW from Chittagong Independent University, Bangladesh. Her research discipline is business intelligence, LAW, and Computational thinking. She has done 3 (2020).

[17] NARAYAN, VIPUL, A. K. Daniel, and Pooja Chaturvedi. "FGWOA: An Efficient Heuristic for Cluster Head Selection in WSN using Fuzzy based Grey Wolf Optimization Algorithm." (2022).

[18] Faiz, Mohammad, et al. "IMPROVED HOMOMORPHIC ENCRYPTION FOR SECURITY IN CLOUD USING PARTICLE SWARM OPTIMIZATION." Journal of Pharmaceutical Negative Results (2022): 4761-4771.

[19] Narayan, Vipul, A. K. Daniel, and Pooja Chaturvedi. "E-FEERP: Enhanced Fuzzy based Energy Efficient Routing Protocol for Wireless Sensor Network." Wireless Personal Communications (2023): 1-28.

[20] Tyagi, Lalit Kumar, et al. "Energy Efficient Routing Protocol Using Next Cluster Head Selection Process In Two-Level Hierarchy For Wireless Sensor Network." Journal of Pharmaceutical Negative Results (2023): 665-676.

[21] Paricherla, Mutyalaiah, et al. "Towards Development of Machine Learning Framework for Enhancing Security in Internet of Things." Security and Communication Networks 2022 (2022).
[22] Sawhney, Rahul, et al. "A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease." Decision Analytics Journal 6 (2023): 100169.
[23] Srivastava, Swapnita, et al. "An Ensemble Learning Approach For Chronic Kidney Disease Classification." Journal of Pharmaceutical Negative Results (2022): 2401-2409.
[24] Mall, Pawan Kumar, et al. "FuzzyNet-Based Modelling Smart Traffic System in Smart Cities Using Deep Learning Models." Handbook of Research on Data-Driven Mathematical Modeling in Smart Cities. IGI Global, 2023. 76-95.
[25] Mall, Pawan Kumar, et al. "Early Warning Signs Of Parkinson's Disease Prediction Using Machine Learning Technique." Journal of Pharmaceutical Negative Results (2022): 4784-4792.
[26] Pramanik, Sabyasachi, et al. "A novel approach using steganography and cryptography in business intelligence." Integration Challenges for Analytics, Business Intelligence, and Data Mining. IGI Global, 2021. 192-217.
[27] Narayan, Vipul, et al. "Deep Learning Approaches for Human Gait Recognition: A Review." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023.
[28] Narayan, Vipul, et al. "FuzzyNet: Medical Image Classification based on GLCM Texture Feature." 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). IEEE, 2023