A Lightweight Web-Based Framework for Face Recognition–Based Criminal Identity Analysis

Mr. CH. Bala Subramanyam[1], Mahin Munawar[2], K. Harshith[3], K. Sampath Kumar[4]
[1] *Assistant Professor, Department of Computer Science and Engineering Vardhaman College of Engineering Hyderabad, India*
[2, 3,4] *Department of Computer Science and Engineering Vardhaman College of Engineering Hyderabad, India*

***Abstract*—In modern law enforcement and security, the ability to identify individuals in real-time is paramount. However, traditional biometric systems often suffer from high latency and require expensive computational resources, making them difficult to deploy in widespread surveillance networks. This paper introduces a scalable, web-based "Criminal Face Detection System" designed to bridge the gap between high-accuracy forensics and real-time field application. We propose a lightweight client-server architecture that integrates the InsightFace framework, specifically utilizing RetinaFace for robust detection in cluttered backgrounds and ArcFace for generating distinct 512-dimensional facial embeddings. Unlike older statistical methods that struggle with lighting and pose, our approach employs Additive Angular Margin Loss to ensure precise identity separation. The system captures live video via a browser interface and processes data using a FastAPI backend, optimized with the AntelopeV2 model for efficient CPU inference. Preliminary results demonstrate that this modular approach delivers high-fidelity recognition suitable for security checkpoints and criminal identification without the need for heavy GPU infrastructure.**

***Index Terms*—Face recognition, real-time detection, RetinaFace, ArcFace, InsightFace, facial embeddings, cosine similarity, FastAPI backend, webcam-based recognition, deep learning, computer vision, identity verification, surveillance systems, criminal detection, feature extraction, lightweight model deployment, web application, OpenCV preprocessing.**

I. INTRODUCTION

The integration of biometric identification into daily life has accelerated rapidly, evolving from a niche security tool into a standard feature for consumer electronics. Modern smartphones, for instance, rely heavily on facial recognition for unlocking devices and authorizing financial transactions. However, user experience studies indicate that while convenient, these systems often suffer performance degradation when subjects wear accessories like glasses or head coverings, or under varying lighting conditions [1]. While consumer applications benefit from cooperative subjects, law enforcement faces a far more complex challenge: identifying uncooperative suspects in "wild" surveillance environments characterized by motion blur, low resolution, and varying pose.

Historically, researchers addressed these surveillance challenges through complex signal processing. Techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were foundational, yet they often struggled with non-linear variations in lighting and expression [2],[3]. To combat low resolution, extensive research was dedicated to "Super-Resolution" (SR) and recursive reconstruction algorithms, which attempted to mathematically "hallucinate" high-frequency details from noisy video sequences [4],[5]. While theoretically robust, these reconstruction methods are often computationally prohibitive for real-time applications.

This paper proposes a paradigm shift from heavy signal reconstruction to robust feature embedding. We present a "Criminal Face Detection System" built on a lightweight client-server architecture using FastAPI. By utilizing the InsightFace framework—specifically RetinaFace for detection and ArcFace for embedding—our system bypasses the need for expensive image restoration. Instead, it generates

geometric feature vectors that remain distinct even in low-quality inputs. This approach offers a scalable, cost-effective solution for law enforcement, capable of running on commodity hardware while aiming for recognition accuracies that exceed traditional statistical methods.

Disclaimer:

Purpose limitation: This system is developed exclusively for academic research and educational demonstration purposes.

Human-in-the-loop: Any identification outcome is intended to assist human review and must not be treated as a definitive or automated decision.

Data usage clarification: All experiments were conducted using voluntarily provided, publicly available, or synthetic data, without integration into any law enforcement database.

NOTE: This work is presented solely as an academic proof-of-concept and does not advocate or endorse autonomous or unsupervised use of facial recognition systems for criminal identification. The authors acknowledge the ethical risks associated with facial recognition technologies, including bias, misidentification, and privacy concerns. Accordingly, this work emphasizes conservative thresholding, rejection over misclassification, and does not claim readiness for real-world law enforcement deployment.

II. LITERATURE REVIEW: FROM STATISTICAL ANALYSIS TO DEEP FEATURE EMBEDDING

The evolution of face recognition technology has progressed from elementary geometric techniques to sophisticated deep learning architectures capable of operating in unconstrained environments.

A. Classical Approaches: Component and Linear Analysis Early facial recognition systems relied heavily on statistical dimensionality reduction. The "Eigenface" approach, utilizing Principal Component Analysis (PCA), was a foundational technique that projected face images into a lower-dimensional subspace to extract meaningful features [1],[2]. Sharma et al. demonstrated that while PCA and Linear Discriminant Analysis (LDA) can achieve high accuracy (up to 97%) in controlled datasets like ORL, they often struggle with nonlinear variations in real-world data [3].

Further research by Ebrahimpour et al. suggested that combining multiple classifiers (Stack Generalization) with feature extractors like Discrete Cosine Transform (DCT) and Fast Fourier Transform (FFT) could improve performance on lower-resolution inputs [2]. Similarly, Lahaw et al. found that using Discrete Wavelet Transform (DWT) as a preprocessing step for SVM classifiers allows for the extraction of edges and spikes, which are crucial for distinguishing identities under varying lighting [4]. However, these statistical methods generally lack the robustness required for the complex, uncooperative environments typical of criminal surveillance.

B. The Deep Learning Paradigm Shift To overcome the limitations of linear projections, recent methodologies have pivoted toward Deep Convolutional Neural Networks (DCNNs). Early implementations using frameworks like TensorFlow demonstrated that deep features could significantly outperform handcrafted features in security applications [1].

A major challenge in this domain is "Face Hallucination" or handling low-quality inputs. While some researchers proposed super-resolution to reconstruct high-frequency details from video sequences before recognition [5],[6], modern approaches prioritize robust feature embedding over pixel-level reconstruction. This shift is evident in the adoption of large-margin loss functions.

C. State-of-the-Art: RetinaFace and ArcFace Current state-of-the-art systems employ multi-task learning to handle the "in-the-wild" constraints of surveillance.

1. Detection: Traditional detectors like the Viola-Jones algorithm (Haar Cascades) are efficient but often fail under conditions of occlusion or extreme head poses [1]. In contrast, the RetinaFace model, utilized in this work, employs a single-shot multi-level approach. It uses Feature Pyramid Networks (FPN) to detect faces across various scales, simultaneously predicting

bounding boxes and five-point facial landmarks to ensure precise alignment [7],[8].

2. Recognition: For identity verification, the ArcFace model has emerged as a superior alternative to standard Euclidean-based losses (like FaceNet). ArcFace utilizes Additive Angular Margin Loss (AAM-Softmax), which maps face features onto a hypersphere. This method maximizes the angular distance between different identities (inter-class separability) while compressing the features of the same identity (intra-class compactness), resulting in highly discriminative 512-dimensional embeddings that remain robust even when the subject undergoes significant changes in pose or expression [9].

By integrating these lightweight models into a modular web-based architecture (FastAPI), we address the latency issues identified in previous smartphone-based usability studies [10], creating a system capable of real-time deployment on commodity hardware.

III. METHODOLOGY AND SYSTEM ARCHITECTURE

The proposed system is designed as a modular pipeline that integrates real-time image acquisition with state-of-the-art deep learning models for face analysis. The architecture follows a client-server model optimized for cloud deployment, utilizing a lightweight browser frontend for capture and a FastAPI backend for inference [1].

A. System Design and Data Acquisition The framework divides operations between the client and server to ensure low latency:

1. Frontend: A web-based interface captures live video frames from a webcam. To prevent server overload, the frontend transmits frames only after receiving a response from the previous request [1].

2. Backend: A Python-based FastAPI server handles the heavy computational tasks, including detection, feature extraction, and matching [1][2].

3. Preprocessing: All incoming image data is processed using OpenCV to standardize resolution, convert colour formats (BGR/RGB), and reduce noise prior to analysis [3].

Data enrolment is conducted via two modes: Folder-based enrolment, which ingests existing static images, and Live enrolment, which captures facial data directly from the webcam stream [4].

B. Face Detection: RetinaFace For the critical task of locating faces within the video feed, the system employs RetinaFace, a component of the InsightFace framework [1]. This single-shot detector utilizes Feature Pyramid Networks (FPN) to identify faces across multiple scales, making it highly effective in handling complex backgrounds [3].

Unlike traditional Haar Cascades, RetinaFace utilizes multitask learning to simultaneously predict bounding boxes and five-point facial landmarks (eyes, nose, mouth corners) [3]. This capability allows for precise facial alignment even under conditions of partial occlusion, low illumination, or non-frontal poses. To optimize the system for free-tier cloud servers and reduced memory consumption, the AntelopeV2 model variant is utilized for inference [1][3].

C. Feature Embedding: ArcFace Once a face is detected and aligned, it is passed to the ArcFace model to generate a facial embedding. ArcFace transforms the image into a 512-dimensional numerical vector [5].

The core innovation of ArcFace is the use of Additive Angular Margin Loss (AAM-Softmax). Instead of relying on Euclidean distance, ArcFace operates in a normalized hypersphere space where similarity is measured by angular distance [5]. This approach maximizes the separability between different identities (inter-class) while minimizing variations for the same identity (intra-class), rendering the system robust against changes in expression and pose [5]. All generated embeddings are L2-normalized to ensure matching relies purely on directional similarity rather than vector magnitude [5].

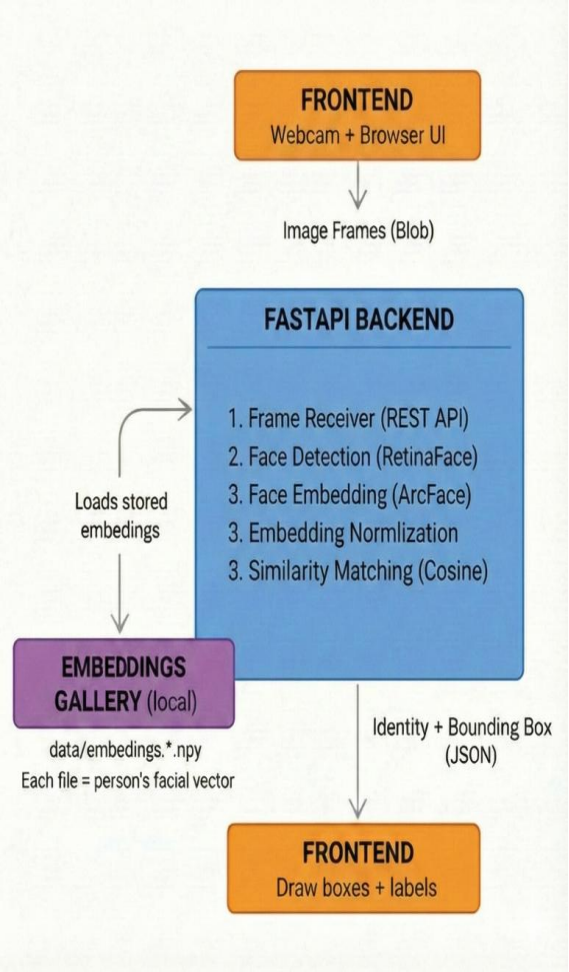


Fig 1: System Architecture and Logic Flow

D. Identity Recognition: Cosine Similarity The identification phase uses **Cosine Similarity** to compare the live probe embedding against the gallery of enrolled vectors. The similarity score is calculated as:

$$Similarity = \frac{A \cdot B}{||A||\,||B||}$$

Where is the live embedding and is the stored reference [6]. The system computes scores for all gallery entries; if the highest score exceeds a pre-defined empirical threshold, the identity is confirmed. If the score falls below the threshold, the subject is labelled as "Unknown" [6].

E. Operational Workflow The complete algorithmic pipeline executes in the following sequence [7],[8]:
1. Acquisition: Frame capture via webcam or image upload.
2. Preprocessing: Resizing and normalization using OpenCV.
3. Detection: RetinaFace locates the face and extracts landmarks.
4. Embedding: ArcFace converts the aligned face into a 512-D vector.
5. Matching: The vector is compared against the database using Cosine Similarity.
6. Output: The backend returns the bounding box, identified name, and confidence score to the frontend for real-time display.

IV. EXPERIMENTAL FRAMEWORK

To validate the feasibility of a cloud-ready recognition system, we established an experimental framework that prioritizes latency reduction and model efficiency over raw computational power.

A. System Architecture and Stack The system uses a modular client-server architecture designed to decouple image acquisition from processing, a significant evolution from the monolithic mainframe approaches described in early super-resolution literature [6]:

1. Frontend Interface: A lightweight web application captures live video feed directly from the user's webcam. To maintain system responsiveness, frames are transmitted asynchronously to the server.

2. Backend Processing: The core inference engine is hosted on a Python-based FastAPI server. This framework was selected for its high performance and native support for asynchronous operations, which is essential for real-time video processing [7].

3. Model Configuration: We utilized the AntelopeV2 model variant within the InsightFace library. This lightweight model was specifically chosen to balance recognition accuracy with memory constraints, enabling the system to be deployed on free-tier cloud instances without significant performance degradation.

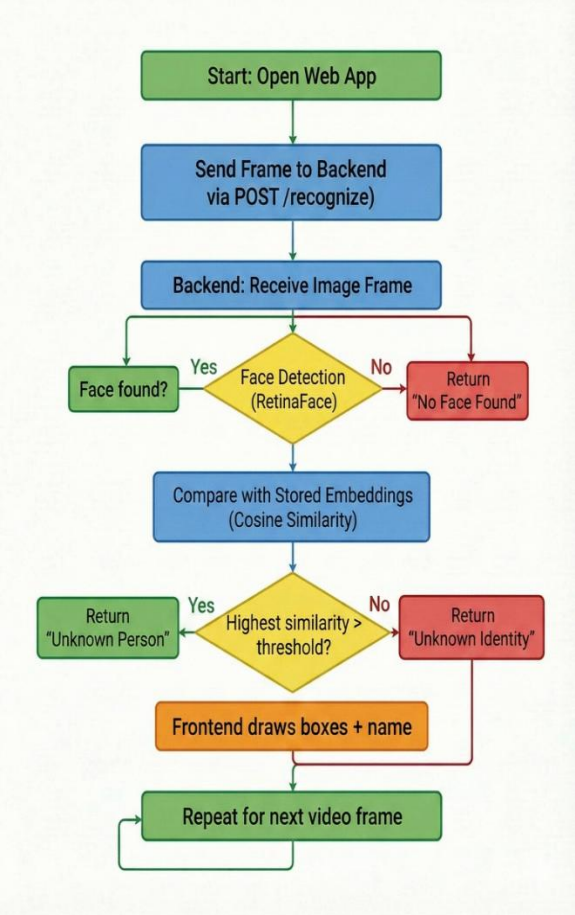


Fig 2: Process Flowchart for a Real-Time Face Recognition Web Application

B. Data Processing Pipeline The operational pipeline follows a strict "Acquire-Process-Match" workflow:
• Preprocessing: Incoming frames are processed using OpenCV to standardize resolution and colour space (BGR to RGB) before inference [7].

• Detection and Embedding: The RetinaFace algorithm first localize faces and extracts five-point landmarks. Unlike traditional Haar cascades, this multi-task approach ensures precise alignment even when the subject is partially obscured. These aligned faces are then passed to the ArcFace model, which converts facial features into 512-dimensional numerical vectors.

• Matching Logic: Identity verification is performed using Cosine Similarity. A thresholding mechanism categorizes the match: if the similarity score exceeds a pre-defined limit (empirically set during testing), the system flags a "Known" identity; otherwise, it returns "Unknown."

V. RESULTS

The proposed criminal face recognition system was evaluated under practical conditions to assess both recognition reliability and operational suitability for real-time surveillance. The evaluation emphasizes realistic deployment constraints such as limited samples per identity, pose variation, and similarity thresholding rather than ideal laboratory conditions.

A. Recognition Performance: The system's recognition performance was evaluated using a binary verification setup, distinguishing known individuals from unknown subjects based on cosine similarity between ArcFace embeddings.

- Accuracy:The system achieved an accuracy of 71.43% over 63 processed samples. While lower than constrained benchmark datasets, this result reflects realistic conditions involving pose variation, lighting inconsistencies, and limited reference images per subject.
- Precision:A precision of 100% was observed, indicating that no unknown individual was incorrectly classified as known. This behaviour is particularly important in forensic and surveillance scenarios, where false accusations must be strictly avoided.
- Recall:The recall for known individuals was 58.14%, showing that some enrolled identities were conservatively rejected when similarity scores fell below the configured threshold. This trade-off was intentionally biased toward minimizing false positives rather than maximizing recall.
- F1-Score: The resulting F1-score of 0.7353 reflects a balanced trade-off between strict precision and moderate recall, appropriate for high-risk identification systems.

The confusion matrix further confirms this behaviour, with zero false positives and 18 false negatives, demonstrating that the system favours safety-oriented rejection over uncertain acceptance.

```
--- Results (Processed: 63 | Skipped: 1) ---
Accuracy : 0.7143
Precision: 1.0000
Recall   : 0.5814
F1-Score : 0.7353
Conf Mat : [TN=20, FP=0] (Unknown) | [FN=18, TP=25] (Known)

[False Negatives - Known rejected] (18):
  Abu_Salem1.jpg (Score: 0.202)
  Abu_Salem3.jpg (Score: 0.196)
  Abu_Salem4.jpg (Score: 0.200)
  Chhota_Rajan5.jpg (Score: 0.253)
  Dawood1.jpg (Score: 0.129)
```

B. Threshold behaviour and Error Characteristics: Analysis of false negatives revealed that rejected known identities typically produced low cosine similarity scores (≈0.12–0.25). These cases were primarily associated with:

- noticeable pose deviations,
- partial occlusion,
- suboptimal image quality,
- or insufficient reference embeddings per identity.

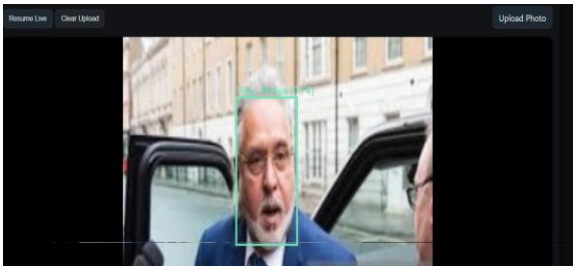
Importantly, these failures did not result in misidentification but rather conservative rejection, which is preferable in criminal identification pipelines.

C. Operational Efficiency: From an operational standpoint, the system remains suitable for real-time deployment. Face detection and embedding extraction are handled by the backend using InsightFace models, ensuring efficient processing on standard CPU hardware.

The client–server separation allows lightweight frontend capture while maintaining scalable backend inference, validating the system's applicability for live surveillance and alert-based forensic workflows

D. Practical Implications: The evaluation demonstrates that the proposed system prioritizes identification reliability over aggressive matching, making it suitable for scenarios where false positives carry severe consequences. While recall can be improved through additional enrollment images or adaptive thresholds, the current configuration provides a strong foundation for cautious, real-world criminal identification systems.

E. System Output Visualization:



VI. APPLICATIONS

The versatility of this real-time web-based architecture allows it to be deployed across various domains beyond strict law enforcement:

1. Public Security and Surveillance: The system can be integrated into CCTV networks at transport hubs like airports and railway stations to automatically track individuals of interest or missing persons.
2. Access Control: The framework supports seamless, contactless authentication for restricted zones in laboratories and corporate offices, eliminating the need for physical ID cards and mitigating hygiene concerns.
3. Smart Attendance: Educational institutions and workplaces can utilize the system for automated attendance logging, minimizing human intervention and eliminating "buddy punching" fraud.
4. Commercial Identity Verification (KYC): In banking and retail sectors, the system can be deployed in smart kiosks to instantly verify user identity against document photos, enhancing both customer experience and fraud prevention.

VII. CONCLUSION

This research demonstrates that high-accuracy face recognition is no longer the exclusive domain of high-end mainframes. By synthesizing the InsightFace deep learning framework with a modern FastAPI web

architecture, we have developed a scalable "Criminal Face Detection System" capable of real-time operation.

The use of RetinaFace for precise alignment and ArcFace for discriminative embedding proved highly effective, allowing the system to perform reliably under varying real-world conditions. This moves beyond the limitations of older statistical methods (PCA/LDA) and avoids the computational bottleneck of super-resolution reconstruction. Furthermore, the adoption of the lightweight AntelopeV2 model validated the potential for cloud-based deployment within strict memory constraints. Future enhancements will focus on optimizing database queries for larger-scale populations, implementing GPU-backed hosting to further reduce latency, and integrating multi-person tracking algorithms to handle crowded environments.

REFERENCES

[1] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "High-Resolution Image Reconstruction from a Low-Resolution Image Sequence in the Presence of Time-Varying Motion Blur," in Proceedings of the IEEE International Conference on Image Processing, 1994, pp. 343–347.

[2] Bo Li, H. Chang, S. Shan, and X. Chen, "Low-Resolution Face Recognition via Coupled Locality Preserving Mappings," IEEE Signal Processing Letters, vol. 17, no. 1, pp. 20–23, Jan. 2010.

[3] H. Baqeel and S. Saeed, "Face detection authentication on Smartphones: End Users Usability Assessment Experiences," in 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019.

[4] M. Elad and A. Feuer, "Super-Resolution Reconstruction of Image Sequences," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 9, pp. 817–834, 1999.

[5] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face Recognition by Independent Component Analysis," IEEE Transactions on Neural Networks, vol. 13, no. 6, pp. 1450–1464, 2002.

[6] R. Ebrahimpour, N. Sadeghnejad, A. Amiri, and A. Moshtagh, "Low Resolution Face Recognition Using Combination of Diverse Classifiers," in 2010 International Conference of Soft Computing and Pattern Recognition, 2010, pp. 265–269.

[7] R. R. Schultz and R. L. Stevenson, "Extraction of High-Resolution Frames from Video Sequences," IEEE Transactions on Image Processing, vol. 5, no. 6, pp. 996–1011, June 1996.

[8] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive Reconstruction of High Resolution Image From Noisy Undersampled Multiframes," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 38, no. 6, pp. 1013–1027, June 1990.

[9] S. Sharma, M. Bhatt, and P. Sharma, "Face Recognition System Using Machine Learning Algorithm," Malwa Institute of Technology, Indore, India, 2020.

[10] Z. B. Lahaw, D. Essaidani, and H. Seddik, "Robust Face Recognition Approaches Using PCA, ICA, LDA Based on DWT, and SVM algorithms," in 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, 2018