Real-Time Employee Face Recognition System

Authors: Shubhpreet, Jay Mewada

Abstract

This paper presents a real-time face recognition system designed to identify specific employees while distinguishing unknown faces. The proposed pipeline integrates state-of-the-art models such as YOLOv8 for object detection, MTCNN for face detection, FaceNet for embedding generation, and a lightweight neural network for classification. The system achieves an accuracy of 94.74% on test data, demonstrating robustness to pose, lighting, and expression variations. Our method ensures scalability, privacy, and real-time performance, making it suitable for deployment in labor-intensive environments.

1. Introduction

Real-time face recognition is crucial for employee identification and productivity monitoring in labor-intensive sectors. Existing systems often struggle with pose variations, inconsistent lighting, and real-time constraints. To address these, we propose a modular pipeline that integrates advanced detection and recognition techniques for robust performance.

Although embedding-based methods like FaceNet offer high accuracy, their computational demands can hinder real-time scalability. By embedding FaceNet within a modular pipeline optimized for resource efficiency, we ensure real-time performance without compromising accuracy, bridging the gap between theory and application.

Key Contributions:

- Introduced a modular, scalable pipeline for real-time employee recognition.
- Successfully integrated YOLOv8, MTCNN, and FaceNet for robust performance.
- Ensured privacy by relying on embeddings instead of raw image storage.

2. Related Work

Face recognition research has advanced significantly, with models like YOLO revolutionizing real-time object detection through state-of-the-art speed and accuracy. However, their application to specific domains, such as employee tracking, remains underexplored. Similarly, face detection models like MTCNN excel at landmark detection but often face scalability issues under diverse lighting and pose conditions.

Embedding-based approaches such as FaceNet are robust, converting faces into compact feature representations. However, they depend heavily on diverse datasets, which are challenging to obtain in domain-specific use cases. Moreover, previous works lack the integration of comprehensive pipelines addressing real-time constraints, domain-specific challenges, and privacy concerns.

Unlike models like RetinaFace, which excel at high-precision detection but struggle with real-time deployment, MTCNN offers an effective balance of speed and robustness, making it ideal for scalable applications. Similarly, FaceNet's triplet loss-based embeddings outperform earlier methods by eliminating intermediate steps like PCA, enhancing both efficiency and accuracy.

3. Methodology

3.1 Motivation

General-purpose recognition systems often fail to meet the demands of labor-intensive environments due to limited scalability for adding new employees, inadequate handling of variations in pose, lighting, and occlusions, and the lack of privacy-preserving embeddings for deployment. To address these challenges, we designed a pipeline that emphasizes robustness, efficiency, and modularity, ensuring adaptability to real-world conditions.

3.2 Data Collection

A systematic image collection strategy ensured dataset diversity, covering 63 unique configurations per individual, including 7 poses, 3 lighting setups, and 3 expressions. Accessories like glasses and masks further improved robustness. Data augmentation simulated variations in brightness, contrast, and orientation, enhancing the dataset's real-world applicability.

3.3 System Pipeline

The pipeline for the face recognition system consists of four sequential steps, as illustrated in Figure 1. Each step plays a critical role in processing input data and producing accurate results.

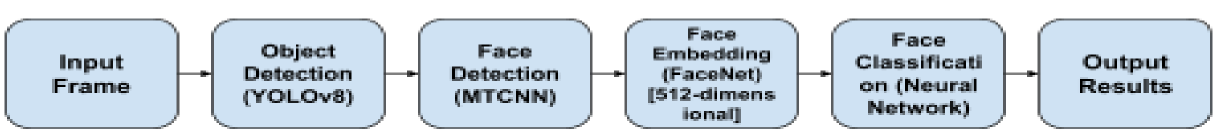

Figure 1: Pipeline Flow Diagram

Step 1: Object Detection (YOLOv8): Detects objects like persons and cups (objects of interest) in video frames, efficiently filtering regions of interest to reduce computational overhead for subsequent stages.

Step 2: Face Detection (MTCNN): Precisely extracts faces from detected regions (persons), ensuring robustness against variations in pose, lighting, and occlusions.

Step 3: Face Embedding (FaceNet): Converts faces into 512-dimensional embeddings for efficient and privacy-preserving feature comparisons.

Step 4: Face Classification (Neural Network)

A lightweight neural network classifies embeddings into specific identities or unknown.

- **Input Layer**: Accepts 512-dimensional embeddings from FaceNet.
- **Hidden Layers**: Two layers (256 and 128 neurons) with ReLU activation for non-linearity and feature extraction.
- **Output Layer**: Uses softmax activation to classify into three classes: two known individuals and unknown.

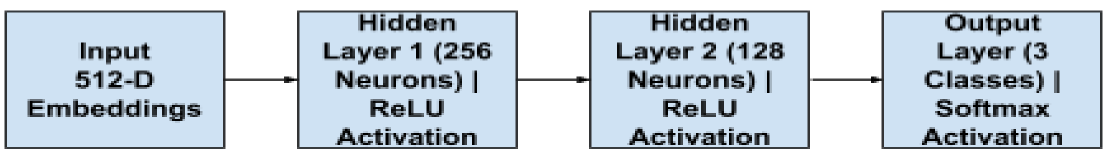

Figure 2: Our Neural Network Architecture

4. Results and Discussion

4.1 Model Performance and Training

Figure 3 illustrates the system's performance, combining the training curve and confusion matrix. The **training curve** shows a steady loss reduction over 50 epochs, indicating effective learning,

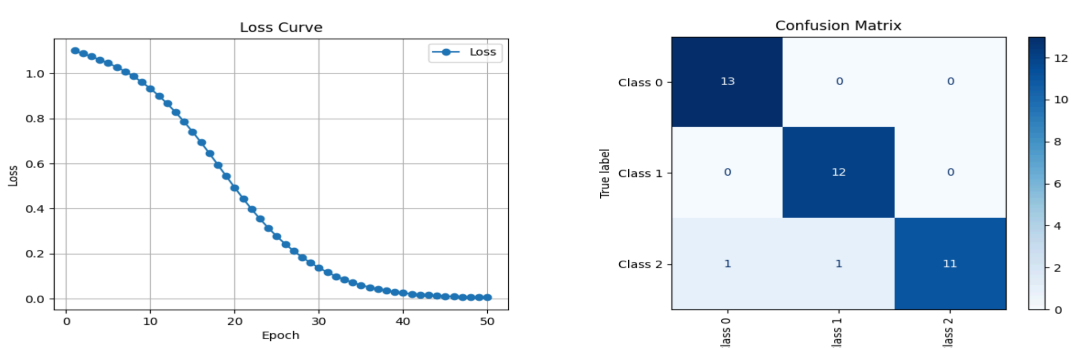

Figure 3: Training Curve and Confusion Matrix

while the **confusion matrix** summarizes classification results across three classes (known individuals [class 0&1] and unknown faces[class 2]), with accurate classification and minimal misclassification.

4.2 Comparative Analysis and Extensibility

Our modular design outperforms traditional systems by enabling scalability and adaptability. Adding new employees requires augmenting the dataset with diverse examples for new classes and retraining the neural network. This extensibility ensures long-term usability in dynamic environments.

4.3 Challenges and Insights

Pose variation, lighting conditions, and handling unknown faces are critical challenges addressed through systematic data collection and augmentation. Incorporating additional unknown examples remains a priority for future improvements.

5. Conclusion and Future Work

This work presents a scalable and privacy-preserving real-time face recognition system that effectively identifies known individuals while handling unknown faces. Its modular design bridges the gap between research advancements and practical applications. Future efforts will focus on expanding the training dataset, enriching the unknown class with diverse examples, and deploying the system in real-world environments, such as manufacturing plants or retail chains, to validate its generalization and track employee productivity more effectively

References

1. Redmon, J., & Farhadi, A. (n.d.). **YOLOv8 Documentation**. Retrieved from https://github.com/ultralytics/ultralytics/blob/main/docs/en/models/yolov8.md
2. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (n.d.). **Joint Face Detection and Alignment using MTCNN**. Retrieved from https://github.com/timesler/facenet-pytorch
3. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). **FaceNet: A Unified Embedding for Face Recognition and Clustering.** *arXiv preprint arXiv:1503.03832*. Retrieved from https://arxiv.org/abs/1503.03832
4. PyTorch Documentation. (n.d.). **PyTorch Framework**. Retrieved from https://pytorch.org
5. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). **Densely Connected Convolutional Networks (DenseNet).** Retrieved from https://arxiv.org/abs/1608.06993