

Implementation of Convolutional Neural Network for Real-Time Face Recognizing Using Python

Muhammad Aditya Rahman¹⁾, Wahyuni²⁾, Pitrasacha Adytia³⁾

^{1,2,3}Teknik Informatika, STMIK Widya Cipta Dharma

Jl. M. Yamin No. 25, Samarinda, Kalimantan Timur

E-mail: 2443902@wicida.ac.id¹⁾, wahyuni@wicida.ac.id²⁾, pitra@wicida.ac.id

ABSTRACT

Workplace security supervision and employee-presence monitoring still depend heavily on direct observation and manual coordination. This study develops a real-time face-recognition system to support individual identification and presence monitoring in the workplace. The system was developed using the Cross-Industry Standard Process for Data Mining (CRISP-DM), covering business understanding, data understanding, data preparation, modeling, evaluation, and deployment. YOLOv8n-Face was used to detect one or multiple faces, while FaceNet with the InceptionResnetV1 architecture pretrained on VGGFace2 generated 512-dimensional facial embeddings. Facial images were processed through square cropping, a five-percent margin, resizing to 160×160 pixels, Contrast Limited Adaptive Histogram Equalization, light sharpening, BGR-to-RGB conversion, pixel standardization, and L2 normalization. Identity matching was performed using cosine similarity against each identity centroid. A similarity threshold and minimum similarity margin were applied to distinguish registered identities from unregistered faces, while Intersection over Union-based tracking and temporal smoothing stabilized predictions across consecutive frames. Testing on independent known and unknown facial images produced a Known Recognition Rate of 87.79%, an Unknown Detection Rate of 100%, a False Acceptance Rate of 0%, a False Rejection Rate of 12.20%, and an overall accuracy of 89.12%. The average processing time was 80.30 milliseconds, equivalent to approximately 12.45 frames per second. These results indicate that the integrated pipeline can recognize multiple registered faces and reject unregistered faces in real time using internal or external cameras.

Keywords: face recognition, YOLOv8n-Face, FaceNet, face embedding, cosine similarity, real-time.

Implementasi Convolutional Neural Network untuk Real-Time Face Recognizing Menggunakan Python

ABSTRAK

Pengawasan keamanan dan pemantauan keberadaan pegawai di lingkungan kerja masih banyak bergantung pada pengamatan langsung dan koordinasi manual. Penelitian ini mengembangkan sistem pengenalan wajah secara *real-time* untuk membantu identifikasi individu dan pemantauan keberadaan pegawai. Pengembangan sistem menggunakan tahapan *Cross-Industry Standard Process for Data Mining* (CRISP-DM), meliputi *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. YOLOv8n-Face digunakan untuk mendeteksi satu atau beberapa wajah, sedangkan FaceNet dengan arsitektur InceptionResnetV1 berbobot *pretrained* VGGFace2 menghasilkan face embedding berdimensi 512. Citra wajah diproses melalui *square cropping*, penambahan margin lima persen, *resize* 160×160 piksel, *Contrast Limited Adaptive Histogram Equalization*, *sharpening* ringan, konversi BGR ke RGB, standarisasi piksel, dan normalisasi L2. Pencocokan identitas dilakukan menggunakan *cosine similarity* terhadap *centroid* setiap identitas. *Similarity threshold* dan minimum *similarity margin* diterapkan untuk membedakan identitas terdaftar dan wajah tidak terdaftar, sedangkan pelacakan berbasis *Intersection over Union* dan *temporal smoothing* menjaga kestabilan prediksi antar-frame. Pengujian pada data wajah dikenal dan tidak dikenal yang independen menghasilkan *Known Recognition Rate* 87,79%, *Unknown Detection Rate* 100%, *False Acceptance Rate* 0%, *False Rejection Rate* 12,20%, dan *accuracy* 89,12%. Waktu pemrosesan rata-rata sebesar 80,30 milidetik atau sekitar 12,45 *frame per second*. Hasil tersebut menunjukkan bahwa *pipeline* terintegrasi mampu mengenali beberapa wajah terdaftar serta menolak wajah tidak terdaftar secara *real-time* melalui kamera internal maupun eksternal.

Kata Kunci: pengenalan wajah, YOLOv8n-Face, FaceNet, *face embedding*, *cosine similarity*, *real-time*.

1. PENDAHULUAN

Closed Circuit Television (CCTV) telah digunakan secara luas untuk pengawasan visual, tetapi proses identifikasi individu pada rekaman atau tayangan langsung masih bergantung pada petugas. Kondisi ini dapat menyebabkan keterlambatan ketika informasi mengenai keberadaan individu dibutuhkan secara cepat. Permasalahan serupa ditemukan di BPK Perwakilan Provinsi Kalimantan Timur, yang memiliki dua akses utama dan area kantor cukup luas, sementara jumlah

personel keamanan terbatas. Pengecekan keberadaan pimpinan atau pegawai juga masih dilakukan dengan bertanya, menghubungi individu terkait, atau mendatangi ruangan secara langsung.

Pengenalan wajah menyediakan kemampuan identifikasi otomatis berdasarkan karakteristik biometrik wajah. Sistem modern umumnya terdiri atas deteksi wajah, preprocessing, ekstraksi fitur, pencocokan, dan penentuan identitas. Deep face recognition mengubah citra wajah menjadi representasi numerik yang dapat dibandingkan

untuk menentukan kemiripan antar citra (Wang & Deng, 2021). Kecepatan detektor dan konsistensi *preprocessing* menjadi faktor penting karena sistem harus memproses *frame* secara berkelanjutan.

Setiawan dkk. (2025) mengintegrasikan YOLOv8 dan CNN serta memperoleh akurasi pengujian 93,8%, namun performa menurun pada pencahayaan rendah dan oklusi. Terampe & Arif Pramudwiatmoko (2025) menggunakan YOLOv8m, ArcFace, SVM, CLAHE, dan *sharpening* dengan akurasi 93,7% serta kecepatan 7,2 FPS. Nurlita dkk. (2024) menunjukkan bahwa ArcFace memberikan kinerja lebih baik daripada Dlib setelah deteksi menggunakan YOLOv8. Penelitian tersebut memperlihatkan efektivitas *pipeline* deteksi dan pengenalan, tetapi sebagian besar menggunakan klasifikator tertutup yang kurang fleksibel dalam menolak wajah asing atau menambah identitas baru.

Kebaruan penelitian ini terletak pada integrasi YOLOv8n-Face, FaceNet, *centroid* identitas, *cosine similarity*, *similarity threshold*, *similarity margin*, pelacakan berbasis *Intersection over Union* (IoU), dan *temporal smoothing* dalam satu *pipeline* multiwajah. Identitas baru dapat ditambahkan melalui pembaruan embedding dan centroid tanpa melatih ulang FaceNet. Sistem juga secara eksplisit mengevaluasi pengenalan wajah dikenal dan penolakan wajah tidak dikenal melalui *Known Recognition Rate* (KRR), *Unknown Detection Rate* (UDR), *False Acceptance Rate* (FAR), dan *False Rejection Rate* (FRR).

1.1. Kondisi Pengawasan dan Kebutuhan Sistem

Pada lingkungan kantor, kebutuhan pengawasan tidak hanya berkaitan dengan pencatatan kejadian, tetapi juga dengan kemampuan memperoleh informasi mengenai siapa yang berada pada suatu area. CCTV konvensional mampu menampilkan dan merekam aktivitas, namun tidak secara otomatis mengubah informasi visual tersebut menjadi identitas individu. Akibatnya, petugas tetap harus mengamati tayangan kamera, menelusuri rekaman, atau mengonfirmasi keberadaan seseorang secara langsung. Mekanisme tersebut kurang efisien ketika informasi dibutuhkan segera dan jumlah titik pengawasan lebih banyak daripada personel yang tersedia.

Kondisi serupa ditemukan pada BPK Perwakilan Provinsi Kalimantan Timur. Area kantor memiliki akses depan yang dilengkapi resepsionis dan petugas keamanan serta akses belakang yang menggunakan kunci akses tanpa penjagaan langsung. Luas area dan keterbatasan jumlah personel menyebabkan tidak seluruh titik dapat diawasi secara intensif pada waktu yang sama. Selain aspek keamanan, pegawai juga memerlukan cara yang lebih cepat untuk mengetahui keberadaan pimpinan atau rekan kerja tanpa harus mendatangi ruangan, bertanya kepada pegawai lain, atau menghubungi orang yang bersangkutan.

Sistem yang dikembangkan pada penelitian ini tidak dimaksudkan untuk menggantikan fungsi petugas keamanan. Sistem diposisikan sebagai prototipe alat bantu identifikasi yang mengolah citra kamera secara *real-time*,

menampilkan lokasi wajah dalam bentuk *bounding box*, memberikan label identitas untuk wajah yang terdaftar, dan memberikan label *Unknown* untuk wajah yang tidak memenuhi kriteria penerimaan. Informasi tersebut diharapkan dapat mempercepat proses verifikasi visual dan mendukung pemantauan keberadaan individu secara lebih terstruktur.

1.2. State of the Art, Kesenjangan, dan Kontribusi

Penelitian Setiawan dkk. (2025) memperlihatkan bahwa kombinasi YOLOv8 dan CNN dapat digunakan untuk pengenalan wajah real-time dengan akurasi pengujian 93,8% dan F1-score 0,93. Namun, performa sistem menurun pada pencahayaan rendah dan oklusi. Terampe & Arif Pramudwiatmoko (2025) menggunakan YOLOv8m, ArcFace, dan SVM pada sistem presensi tanpa kontak serta memperoleh akurasi 93,7% dan kecepatan 7,2 FPS. Penelitian tersebut juga menunjukkan pentingnya CLAHE, *denoising*, dan *sharpening* untuk memperbaiki kualitas citra sebelum proses pengenalan.

Nurlita dkk. (2024) membandingkan ArcFace dan Dlib setelah deteksi menggunakan YOLOv8. ArcFace menghasilkan *accuracy* 0,90 dan F1-score 0,95, sedangkan Dlib menghasilkan *accuracy* 0,57 dan F1-score 0,73. Perbedaan tersebut menunjukkan bahwa metode representasi wajah setelah tahap deteksi mempunyai pengaruh besar terhadap kualitas pengenalan. Ketiga penelitian terdahulu memperkuat penggunaan detektor satu tahap yang cepat dan model embedding berukuran tetap, tetapi belum secara bersamaan menekankan kalibrasi label *Unknown*, *margin* antar identitas, *centroid*, *tracking* sederhana, *temporal smoothing*, serta penggunaan kamera internal dan eksternal.

Kontribusi penelitian ini adalah menyusun *pipeline* yang menggabungkan YOLOv8n-Face, *preprocessing* citra, FaceNet InceptionResnetV1 *pretrained* VGGFace2, normalisasi L2, *centroid* identitas, *cosine similarity*, *similarity threshold*, dan *similarity margin*. Sistem juga menambahkan pelacakan berbasis *Intersection over Union* dan *temporal smoothing* agar label tidak mudah berubah pada setiap frame. Pemisahan *data enrollment*, validasi, dan pengujian dari sesi pengambilan yang berbeda digunakan untuk mengurangi bias akibat kemiripan latar belakang, pakaian, atau kondisi pengambilan data.

Tabel 1. Perbandingan Penelitian Terdahulu

Table 1. Comparison of Previous Studies

Penelitian	Deteksi	Pengenalan	Hasil Utama	Kesenjangan
Setiawan dkk. (2025)	YOLO v8	CNN <i>transfer learning</i>	Akurasi 93,8%; F1 0,93	Menurun pada cahaya rendah dan oklusi
Terampe & Arif Pramudwiatmoko (2025)	YOLO v8m	ArcFace + SVM	Akurasi 93,7%; 7,2 FPS	Fokus presensi dan klasifikasi SVM
Nurlita dkk. (2024)	YOLO v8	ArcFace/Dlib	ArcFace F1 0,95	Belum membahas kalibrasi <i>Unknown</i>

Penelitian	Deteksi	Pengenalan	Hasil Utama	Kesenjangan
Penelitian ini	YOLO v8n-Face	FaceNet + centroid + cosine similarity	Accuracy 89,12%; UDR 100%; FAR 0%	Prototipe terbatas dan tanpa <i>anti-spoofing</i>

2. RUANG LINGKUP

Penelitian mencakup pengembangan prototipe pengenalan wajah *real-time* pada lingkungan kerja. Dataset diperoleh melalui pengambilan gambar langsung dengan variasi pose depan, kiri, kanan, atas, bawah, dan ekspresi. Data dipisahkan menurut fungsi menjadi *enrollment*, *validation*, dan test dari sesi pengambilan yang berbeda.

Sistem dibatasi pada deteksi wajah menggunakan YOLOv8n-Face, *preprocessing* citra 160×160 piksel, ekstraksi embedding FaceNet, pencocokan *centroid* dengan *cosine similarity*, penolakan wajah asing, *multi-face tracking* sederhana, dan input kamera lokal. Penelitian tidak mencakup aplikasi seluler, *cloud*, integrasi penuh jaringan CCTV, basis data monitoring terpusat, maupun *face anti-spoofing*.

Hasil yang ditargetkan adalah sistem yang mampu mendeteksi serta mengenali beberapa wajah terdaftar, memberikan label *Unknown* pada wajah tidak terdaftar, dan mempertahankan kecepatan yang memadai untuk pemrosesan *real-time*.

3. BAHAN DAN METODE

3.1. Pengenalan Wajah dan FaceNet

FaceNet menghasilkan *embedding* wajah pada ruang *Euclidean* sehingga citra dari identitas yang sama cenderung berdekatan dan identitas berbeda berjauhan. Penelitian ini menggunakan InceptionResnetV1 *pretrained* VGGFace2 melalui facenet-pytorch. Setiap citra menghasilkan *embedding* berdimensi 512, kemudian dinormalisasi menggunakan L2 agar perbandingan berfokus pada arah vektor. FaceNet dipilih karena representasi *embedding* memungkinkan penambahan identitas tanpa pelatihan ulang keseluruhan model. (Rachmansyah dkk., 2025)

3.2. Deteksi YOLOv8n-Face

YOLO melakukan deteksi objek dalam satu kali proses inferensi sehingga sesuai untuk kebutuhan *real-time*. YOLOv8n-Face digunakan untuk menghasilkan *bounding box* satu atau beberapa wajah. Varian *nano* dipilih karena memiliki beban komputasi lebih ringan dan dapat dijalankan pada CPU maupun GPU. Perkembangan arsitektur YOLO dan karakteristik efisiensinya dibahas secara komprehensif oleh Terven & Cordova-Esparza (2024).

3.3. Metode CRISP-DM

Pengembangan sistem mengikuti CRISP-DM yang terdiri atas *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Kerangka ini digunakan karena menyediakan alur iteratif

dari identifikasi kebutuhan sampai penerapan prototipe dan evaluasi kinerja. (Rianti dkk., 2023)



Gambar 1. Tahapan Data Preparation
Figure 1. Data Preparation Stages

Pada tahap *data preparation*, *bounding box* hasil deteksi diubah menjadi *square crop*, ditambah *margin* 5%, di-*resize* menjadi 160×160 piksel, diperbaiki kontrasnya menggunakan CLAHE, dan diberi *sharpening* ringan. CLAHE digunakan untuk meningkatkan kontras lokal tanpa memperkuat *noise* secara berlebihan, sedangkan *sharpening* memperjelas tepi fitur wajah. Citra selanjutnya dikonversi dari BGR ke RGB dan distandardisasi sesuai kebutuhan FaceNet.

3.4. Pembentukan Embedding dan Centroid

Seluruh citra *enrollment* diproses untuk menghasilkan *embedding*. *Embedding* dikelompokkan berdasarkan identitas, kemudian nilai rata-ratanya dihitung menjadi *centroid* dan dinormalisasi kembali dengan L2. Untuk sebuah *embedding* masukan x dan *centroid* c , *cosine similarity* dihitung berdasarkan persamaan (1).

$$\text{sim}(x, c) = (x \cdot c) / (|x|_2 |c|_2) \quad (1)$$

Identitas kandidat ditentukan dari centroid dengan nilai kemiripan tertinggi. Prediksi diterima apabila *best similarity* melampaui *threshold* dan selisih terhadap kandidat kedua memenuhi minimum *similarity* margin. Apabila salah satu syarat tidak terpenuhi, wajah diberi label *Unknown*.

3.5. Kalibrasi, Tracking, dan Temporal Smoothing

Similarity threshold dan *minimum similarity margin* ditentukan menggunakan data validasi yang berisi wajah dikenal serta tidak dikenal. Kombinasi parameter diuji untuk memperoleh keseimbangan antara kemampuan mengenali wajah terdaftar dan menolak wajah asing. Pada video, *bounding box* antar-frame dipasangkan berdasarkan IoU. Riwayat label setiap track disimpan dan keputusan akhir diperoleh melalui voting temporal untuk mengurangi perubahan label sesaat.

```

{
  "method": "YOLOv8n-Face + FaceNet + Cosine Similarity",
  "similarity_threshold": 0.7500000000000001,
  "minimum_similarity_margin": 0.08,
  "overall_accuracy": 0.8911917098445595,
  "known_recognition_rate": 0.877906976744186,
  "unknown_detection_rate": 1.0,
  "false_acceptance_rate": 0.0,
  "false_rejection_rate": 0.12209302325581395,
  "known_misidentification_rate": 0.0,
  "face_detection_rate": 1.0,
  "average_processing_time_ms": 80.30956943007533,
  "estimated_fps": 12.45181622933104,
  "known_test_count": 172,
  "unknown_test_count": 21,
  "total_test_count": 193
}

```

Gambar 2. Hasil Kalibrasi Threshold
Figure 2. Threshold Calibration Result

3.6. Skenario Evaluasi

Evaluasi dilakukan pada test data independen. Metrik yang digunakan meliputi *accuracy*, *precision*, *recall*, F1-score, KRR, UDR, FAR, FRR, waktu pemrosesan, dan FPS. Pengujian fungsional dilakukan melalui *black-box testing*, sedangkan penerimaan pengguna dinilai melalui *beta testing* kepada 10 pegawai.

3.7. Data Penelitian dan Pembagian Berdasarkan Fungsi

Dataset diperoleh melalui pengambilan citra wajah secara langsung menggunakan kamera. Setiap identitas direkam dengan variasi wajah menghadap depan, menoleh ke kiri dan kanan, melihat ke atas dan bawah, serta variasi ekspresi. Variasi tersebut diperlukan karena citra *real-time* tidak selalu memperlihatkan wajah dalam kondisi frontal. Setelah pengambilan data, citra diperiksa untuk memastikan wajah terlihat jelas, tidak mengalami blur berat, tidak terpotong, dan tidak memuat gangguan yang menyebabkan identitas sulit dikenali.

Berbeda dari klasifikasi CNN yang dilatih dari awal, pipeline FaceNet pada penelitian ini menggunakan model pretrained untuk menghasilkan embedding. Oleh sebab itu, pembagian data dilakukan berdasarkan fungsi. *Data enrollment* digunakan untuk membangun *database embedding* dan *centroid* identitas. Data validasi berisi wajah terdaftar dan tidak terdaftar untuk menentukan *similarity threshold* dan *minimum similarity margin*. Data pengujian digunakan hanya pada evaluasi akhir untuk mengukur kemampuan sistem mengenali wajah terdaftar dan menolak wajah tidak terdaftar.

Data validasi dan pengujian berasal dari sesi pengambilan yang berbeda dari *data enrollment*. Pemisahan sesi penting agar model tidak hanya dinilai pada citra yang sangat mirip dari sisi pakaian, latar belakang, pencahayaan, dan waktu pengambilan. Dengan pendekatan ini, evaluasi lebih mendekati kondisi penggunaan aktual ketika wajah ditangkap kembali melalui kamera pada waktu yang berbeda.

Tabel 2. Fungsi Setiap Kelompok Data
Table 2. Function of Each Data Group

Kelompok Data	Fungsi	Keluaran
Enrollment	Membangun representasi identitas	Embedding dan centroid per identitas
Validasi <i>known</i>	Mengukur <i>similarity</i> wajah terdaftar	Kandidat <i>threshold</i> dan <i>margin</i>
Validasi <i>unknown</i>	Mengukur <i>similarity</i> wajah tidak terdaftar	Batas penolakan <i>Unknown</i>
Pengujian	Evaluasi akhir independen	<i>Confusion matrix</i> dan metrik performa

3.8. Rincian Tahap Preprocessing

Deteksi awal dilakukan menggunakan YOLOv8n-Face. Pada *dataset enrollment* yang dirancang memuat satu individu, apabila lebih dari satu wajah terdeteksi, sistem memilih *bounding box* dengan luas terbesar sebagai wajah utama. Ketentuan tersebut hanya digunakan saat

memproses dataset. Pada mode *real-time*, seluruh *bounding box* yang memenuhi ambang deteksi diproses satu per satu sehingga sistem dapat mengenali beberapa wajah secara bersamaan.

Bounding box persegi diperlukan agar proporsi citra wajah tetap konsisten. Sistem memperluas sisi yang lebih pendek hingga membentuk *square crop*, kemudian menambahkan margin 5%. *Margin* mencegah area penting seperti dahi, dagu, dan sisi wajah terpotong terlalu rapat. Nilai margin dibuat terbatas agar latar belakang tidak mendominasi representasi wajah. Citra kemudian diubah menjadi 160×160 piksel karena ukuran tersebut sesuai dengan input standar FaceNet InceptionResnetV1 yang digunakan.

CLAHE diterapkan untuk memperbaiki kontras lokal tanpa memperkuat *noise* secara berlebihan pada seluruh citra. Tahap ini membantu ketika sebagian wajah lebih gelap karena arah cahaya tidak merata. *Sharpening* ringan digunakan untuk meningkatkan kejelasan tepi pada area mata, hidung, bibir, dan kontur wajah. Setelah itu, citra dikonversi dari format BGR yang digunakan OpenCV menjadi RGB sesuai urutan kanal yang digunakan model FaceNet. Standardisasi nilai piksel dilakukan agar distribusi input konsisten dengan kebutuhan model pretrained.

Preprocessing tidak dapat sepenuhnya menghilangkan pengaruh kualitas kamera, blur gerak, oklusi, dan pose ekstrem. Oleh karena itu, tahap ini diperlakukan sebagai mekanisme penyeragaman input, bukan sebagai pengganti dataset yang representatif. Kualitas data enrollment tetap menjadi faktor utama karena centroid identitas dibentuk dari embedding citra yang tersedia.

3.9. Pembentukan Embedding, Centroid, dan Keputusan Identitas

FaceNet menghasilkan vektor *embedding* berdimensi 512 untuk setiap citra wajah. Vektor tersebut tidak secara langsung berisi nama identitas, melainkan representasi numerik karakteristik wajah. Normalisasi L2 mengubah panjang vektor menjadi satu sehingga perbandingan lebih berfokus pada arah vektor daripada skala nilainya. *Embedding* yang telah dinormalisasi dikelompokkan berdasarkan label identitas.

Centroid dihitung sebagai nilai rata-rata seluruh embedding pada satu identitas, kemudian kembali dinormalisasi. Penggunaan *centroid* membuat *database* lebih ringkas karena prediksi tidak harus membandingkan citra *query* terhadap seluruh *embedding enrollment*. Namun, *centroid* yang baik tetap memerlukan data yang cukup beragam. Apabila data *enrollment* hanya menggambarkan satu pose atau satu kondisi cahaya, *centroid* dapat kurang mewakili variasi penampilan identitas tersebut.

Pada saat inferensi, *embedding query* dibandingkan terhadap seluruh *centroid* menggunakan *cosine similarity* sebagaimana Persamaan (1). Nilai tertinggi menjadi kandidat identitas pertama, sedangkan nilai tertinggi kedua digunakan untuk menghitung *similarity margin*. Kandidat

hanya diterima ketika nilai *similarity* mencapai *threshold* dan selisih terhadap kandidat kedua mencapai minimum *margin*. Dua kriteria tersebut mencegah sistem menerima prediksi ketika kemiripan absolut rendah atau ketika dua identitas menghasilkan nilai yang terlalu berdekatan.

$$\text{margin} = \text{sim terbaik} - \text{sim terbaik kedua} \quad (2)$$

Keputusan akhir dapat dirumuskan: identitas diterima jika $\text{sim terbaik} \geq \text{threshold}$ dan $\text{margin} \geq \text{minimum margin}$; selain itu, hasil diberikan sebagai *Unknown*. Nilai *threshold* yang terlalu rendah meningkatkan risiko *false acceptance*, sedangkan *threshold* yang terlalu tinggi meningkatkan *false rejection*. Oleh karena itu, parameter ditentukan melalui data validasi, bukan dipilih hanya berdasarkan perkiraan.

3.10. Tracking dan Temporal Smoothing

Pelacakan dilakukan dengan membandingkan *bounding box* pada *frame* saat ini dan *frame* sebelumnya menggunakan *Intersection over Union*. *Bounding box* dengan tingkat tumpang tindih yang memadai diasosiasikan ke track yang sama. Setiap track menyimpan riwayat label dalam jendela beberapa *frame*. Label akhir dipilih berdasarkan suara terbanyak dengan jumlah minimum vote agar perubahan sesaat tidak langsung mengganti identitas yang ditampilkan.

Status *Analyzing* digunakan ketika jumlah bukti belum mencukupi untuk menentukan label stabil. *Track* akan dihapus apabila wajah tidak lagi terdeteksi selama sejumlah *frame* yang telah ditetapkan. Metode ini ringan dan sesuai dengan tujuan prototipe, tetapi tidak sekuat metode *tracking* yang menggunakan fitur gerak atau *appearance*. Ketika dua wajah berdekatan, saling berpotongan, atau bergerak cepat, track ID masih berpotensi bertukar.

3.11. Metrik Evaluasi

Confusion matrix digunakan untuk melihat hubungan kelas aktual dan kelas prediksi. *Accuracy* menunjukkan proporsi seluruh prediksi yang benar. *Precision* mengukur ketepatan prediksi pada suatu kelas, *recall* mengukur kemampuan menemukan seluruh data pada kelas tersebut, dan *F1-score* merupakan rata-rata harmonik *precision* dan *recall*. *Weighted average* memperhitungkan jumlah sampel setiap kelas sehingga kelas dengan sampel lebih banyak memberikan kontribusi lebih besar.

Evaluasi open-set memerlukan metrik tambahan. *Known Recognition Rate* (KRR) adalah proporsi wajah terdaftar yang dikenali dengan identitas benar. *Unknown Detection Rate* (UDR) adalah proporsi wajah tidak terdaftar yang berhasil diberikan label *Unknown*. *False Acceptance Rate* (FAR) adalah proporsi wajah tidak terdaftar yang salah diterima sebagai identitas terdaftar. *False Rejection Rate* (FRR) adalah proporsi wajah terdaftar yang salah ditolak menjadi *Unknown*. Keempat metrik tersebut diperlukan karena *accuracy* saja tidak menunjukkan kualitas mekanisme penolakan wajah asing.

Kecepatan sistem dihitung dari waktu pemrosesan rata-rata per *frame* dan dinyatakan dalam *frame per second*.

Black-box testing digunakan untuk memeriksa fungsi masukan kamera, deteksi satu dan beberapa wajah, pengenalan identitas, penolakan *Unknown*, dan tampilan informasi. *Beta testing* dilakukan kepada 10 pegawai menggunakan skala 1 sampai 5 untuk menilai kemudahan penggunaan, kecepatan, akurasi, tampilan antarmuka, dan stabilitas sistem.

4. PEMBAHASAN

4.1. Konfigurasi Sistem

Tabel 3. Konfigurasi Model

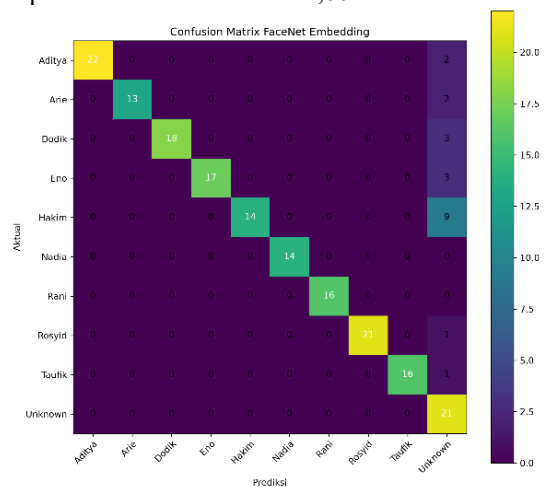
Table 3. Model Configuration

Parameter	Konfigurasi
Detektor wajah	YOLOv8n-Face
Ekstraksi fitur	FaceNet InceptionResnetV1
Bobot model	Pretrained VGGFace2
Ukuran input	160 × 160 × 3 piksel
Dimensi <i>embedding</i>	512
Normalisasi	L2 <i>normalization</i>
Pencocokan	<i>Cosine similarity</i>
<i>Unknown rejection</i>	<i>Similarity threshold</i> dan <i>margin</i>
Pemrosesan	CPU dan GPU

Pipeline menghasilkan berkas *face_embeddings.npz* untuk menyimpan *embedding* dan label, *identity_centroids.npz* untuk menyimpan *centroid*, metadata JSON untuk ringkasan basis data, serta *enrollment_log.csv* untuk mencatat keberhasilan pemrosesan citra.

4.2. Hasil Evaluasi Model

Confusion matrix menunjukkan bahwa tidak terdapat wajah *Unknown* yang diterima sebagai identitas terdaftar. Kesalahan utama berupa wajah terdaftar yang ditolak menjadi *Unknown*, terutama pada identitas dengan variasi kondisi citra yang lebih sulit. Oleh karena itu, *precision* kelas identitas terdaftar mencapai 1,00, sementara *recall* beberapa kelas berada di bawah 1,00.



Gambar 3. Confusion Matrix FaceNet Embedding

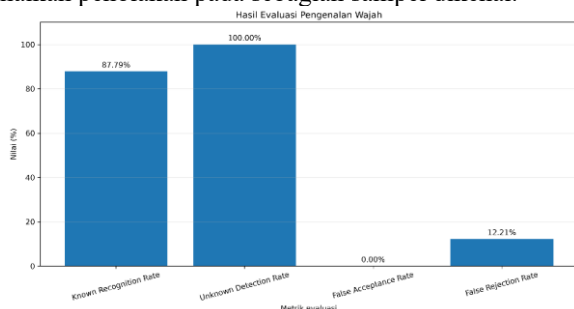
Figure 3. FaceNet Embedding Confusion Matrix

Tabel 4. Hasil Evaluasi Pengenalan Wajah

Table 4. Face Recognition Evaluation Results

Metrik	Hasil
Accuracy	89,12%
Known Recognition Rate	87,79%
Unknown Detection Rate	100%
False Acceptance Rate	0%
False Rejection Rate	12,20%
Kesalahan identifikasi <i>known</i>	0%
Waktu rata-rata	80,30 ms
Kecepatan estimasi	12,45 FPS

Nilai UDR 100% dan FAR 0% menunjukkan seluruh wajah tidak terdaftar berhasil ditolak. KRR 87,79% berarti sebagian besar wajah terdaftar dikenali dengan benar, sedangkan FRR 12,20% menunjukkan masih terdapat wajah terdaftar yang ditolak. Hasil ini mencerminkan konfigurasi yang konservatif: sistem memprioritaskan pencegahan penerimaan wajah asing, dengan konsekuensi kenaikan penolakan pada sebagian sampel dikenal.



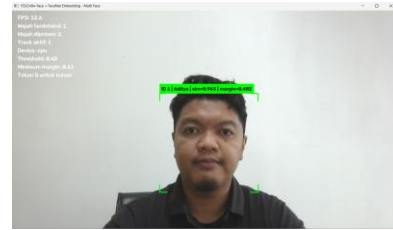
Gambar 4. Ringkasan Evaluasi Pengenalan

Figure 4. Recognition Evaluation Summary

Waktu pemrosesan rata-rata 80,30 milidetik setara dengan sekitar 12,45 FPS. Kecepatan tersebut memungkinkan sistem menampilkan hasil deteksi dan pengenalan secara berkelanjutan, meskipun kelancaran masih dipengaruhi spesifikasi perangkat, resolusi kamera, dan jumlah wajah dalam frame.

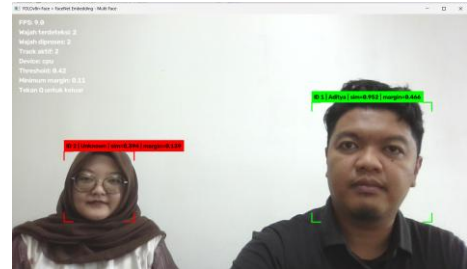
4.3. Pengujian *Real-Time* dan *Deployment*

Pengujian *real-time* mencakup wajah menghadap depan, menoleh kiri dan kanan, melihat atas dan bawah, jarak dekat dan sedang, pencahayaan cukup dan rendah, serta penggunaan aksesoris. Seluruh skenario pada tabel pengujian dinyatakan berhasil. Sistem juga berhasil mendeteksi multiwajah, mengenali identitas terdaftar, memberikan label *Unknown*, dan menampilkan hasil melalui video stream.



Gambar 5. Deployment dengan Deteksi Satu Wajah

Figure 5. Deployment with Single-Face Detection



Gambar 6. Deployment dengan Deteksi Multiwajah

Figure 6. Deployment with Multi-Face Detection

Tampilan sistem memuat *bounding box*, nomor *track*, nama identitas, nilai *cosine similarity*, *similarity margin*, status *Unknown* atau *Analyzing*, jumlah wajah, jumlah *track* aktif, serta FPS. Warna hijau digunakan untuk identitas dikenali, merah untuk *Unknown*, dan kuning untuk proses analisis. *Temporal smoothing* membantu mencegah label berubah akibat satu frame yang kurang baik, sedangkan tracking IoU mempertahankan riwayat prediksi masing-masing wajah.

4.4. Pengujian Fungsional dan Pengguna

Black-box testing terhadap tujuh fungsi utama deteksi wajah, deteksi multiwajah, pengenalan wajah, penolakan wajah tidak dikenal, variasi pencahayaan, variasi posisi, dan pemrosesan *real-time* menunjukkan seluruh fungsi berjalan sesuai keluaran yang diharapkan.

Tabel 5. Hasil Beta Testing pada 10 Pegawai

Table 5. Beta Testing Results from 10 Employees

Aspek	Distribusi Skala 1–5	Rata-rata
Kemudahan penggunaan	0; 0; 2; 3; 5	4,3
Kecepatan sistem	0; 0; 1; 4; 5	4,0
Akurasi pengenalan	0; 0; 1; 6; 3	4,2
Tampilan antarmuka	0; 0; 0; 5; 5	4,5
Stabilitas sistem	0; 0; 1; 5; 4	4,8

Hasil *beta testing* memperlihatkan penilaian dominan pada skala setuju dan sangat setuju. Tampilan antarmuka, kemudahan penggunaan, dan akurasi pengenalan memperoleh respons positif. Nilai pada tabel mengikuti rekapitulasi yang tercantum pada naskah skripsi.

4.5. Analisis Keterbatasan

Sistem masih dipengaruhi pencahayaan sangat rendah, pose ekstrem, jarak terlalu jauh, blur, oklusi, dan resolusi kamera. Tracking IoU bersifat sederhana sehingga track ID dapat berubah ketika beberapa individu bergerak cepat, berpapasan, atau saling menutupi. Kinerja penolakan wajah asing juga bergantung pada variasi *enrollment*, representasi data *Unknown* pada kalibrasi, serta pemilihan *threshold* dan *margin*. Prototipe belum menerapkan *liveness detection* sehingga belum dapat menolak serangan menggunakan foto atau video.

4.6. Hasil Pembentukan Database Embedding

Tahap modeling menghasilkan berkas *face_embeddings.npz* yang memuat label dan *embedding* seluruh citra *enrollment* serta *identity_centroids.npz* yang memuat *labels*, *centroids*, dan *counts*. Struktur tersebut memisahkan data representasi wajah dari kode aplikasi. Ketika identitas baru ditambahkan, proses yang diperlukan adalah menambahkan citra, menjalankan *preprocessing*, menghasilkan *embedding*, dan menghitung ulang *centroid*. Model FaceNet tidak perlu dilatih ulang karena digunakan sebagai ekstraktor fitur *pretrained*.

Pemisahan antara model ekstraksi fitur dan *database* identitas memberikan fleksibilitas dalam pemeliharaan sistem. Meskipun demikian, pembaruan *centroid* perlu dilakukan secara terkendali. Citra baru yang blur, salah label, atau mengandung wajah orang lain dapat menggeser *centroid* dan menurunkan kualitas pengenalan. Oleh sebab itu, tahap verifikasi dataset sebelum pembentukan *embedding* tetap diperlukan.

4.7. Analisis Confusion Matrix dan Classification Report

Data pengujian terdiri atas 172 citra wajah terdaftar dan 21 citra wajah tidak terdaftar, sehingga total evaluasi berjumlah 193 citra. *Weighted precision* mencapai 0,9456, *weighted recall* 0,8912, dan *weighted F1-score* 0,9017. *Recall* yang lebih rendah dibandingkan *precision* sesuai dengan pola kesalahan sistem, yaitu beberapa wajah terdaftar ditolak sebagai *Unknown*, sementara tidak terdapat wajah asing yang diterima sebagai identitas terdaftar.

Confusion matrix memperlihatkan seluruh sampel *Unknown* diprediksi sebagai *Unknown*. Hasil ini menjelaskan UDR 100% dan FAR 0%. Pada kelas *known*, sebagian besar citra berada pada diagonal kelas yang benar, sedangkan kesalahan muncul pada kolom *Unknown*. Tidak ditemukannya kesalahan identifikasi antarkelas *known* menunjukkan bahwa ketika sistem menerima sebuah wajah sebagai identitas terdaftar, identitas yang dipilih sesuai dengan label aktual pada data uji.

Pola tersebut menunjukkan konfigurasi *threshold* cenderung konservatif. Konfigurasi konservatif sesuai untuk konteks pengawasan karena risiko menerima individu asing sebagai pegawai terdaftar dapat dianggap lebih serius daripada meminta verifikasi tambahan terhadap wajah pegawai yang belum dapat dikenali.

Konsekuensinya adalah FRR 12,20%, sehingga sebagian wajah *known* harus diposisikan ulang, diberi pencahayaan lebih baik, atau diamati pada frame berikutnya sebelum memperoleh label stabil.

Tabel 6. Ringkasan Classification Report

Table 6. Classification Report Summary

Ukuran	Nilai	Interpretasi
Jumlah data uji	193 citra	172 <i>known</i> dan 21 <i>unknown</i>
<i>Weighted precision</i>	0,9456	Prediksi yang diterima memiliki ketepatan tinggi
<i>Weighted recall</i>	0,8912	Sebagian <i>known</i> masih ditolak sebagai <i>Unknown</i>
<i>Weighted F1-score</i>	0,9017	Keseimbangan <i>precision</i> dan <i>recall</i> tergolong baik
<i>Accuracy</i>	89,12%	Proporsi seluruh prediksi benar

4.8. Analisis Threshold dan Distribusi Similarity

Grafik distribusi *best cosine similarity* menunjukkan bahwa wajah terdaftar cenderung mempunyai *similarity* lebih tinggi daripada wajah tidak terdaftar. Namun, kedua distribusi dapat memiliki area tumpang tindih. Area ini penting karena menjadi sumber potensial *false acceptance* atau *false rejection*. *Threshold* ditempatkan untuk memisahkan dua kelompok dengan mempertimbangkan kebutuhan keamanan dan tingkat kenyamanan penggunaan.

Similarity margin menambahkan informasi mengenai tingkat kejelasan kandidat. Nilai *similarity* yang tinggi belum selalu cukup apabila kandidat pertama dan kedua hampir sama. Margin kecil menunjukkan bahwa *embedding query* berada dekat dengan lebih dari satu *centroid* sehingga keputusan identitas belum meyakinkan. Dengan mewajibkan *threshold* dan *margin* secara bersamaan, sistem dapat menunda keputusan dan memberi label *Unknown* atau *Analyzing* daripada menampilkan identitas yang meragukan.

Hasil FAR 0% menunjukkan *threshold* dan *margin* berhasil menolak seluruh data *unknown* pada skenario pengujian. Namun, nilai tersebut berlaku pada dataset dan kondisi yang diuji. Penggunaan pada populasi yang lebih besar, kamera berbeda, atau kondisi pencahayaan lebih ekstrem dapat mengubah distribusi *similarity*. Oleh karena itu, *threshold* perlu dikalibrasi ulang ketika komposisi *database* atau lingkungan implementasi berubah secara signifikan.

4.9. Analisis Kecepatan Pemrosesan

Waktu pemrosesan rata-rata 80,30 milidetik menghasilkan estimasi 12,45 FPS. Nilai ini menunjukkan sistem dapat memberikan tampilan bergerak secara berkelanjutan pada perangkat pengujian. Kecepatan dipengaruhi oleh jumlah wajah, resolusi *frame*, perangkat CPU atau GPU, waktu inferensi YOLO, *preprocessing* setiap *bounding box*, inferensi FaceNet, serta proses rendering antarmuka.

Ketika jumlah wajah meningkat, tahap FaceNet dan *preprocessing* dijalankan berulang untuk setiap *bounding box* sehingga beban komputasi bertambah. Untuk menjaga

performa, sistem dapat membatasi resolusi *input*, memilih model deteksi berukuran ringan, atau memproses *embedding* pada *interval frame* tertentu. Temporal *smoothing* membantu strategi tersebut karena label stabil dapat dipertahankan selama beberapa *frame* tanpa harus mengganti hasil pada setiap perubahan kecil.

Nilai 12,45 FPS lebih tinggi daripada 7,2 FPS yang dilaporkan Terampe & Arif Pramudwiatmoko (2025), meskipun perbandingan langsung harus dilakukan secara hati-hati karena perangkat, ukuran model, dataset, resolusi, dan skenario pengujian berbeda. Hasil penelitian ini terutama menunjukkan bahwa *pipeline* dapat dijalankan secara *real-time* pada lingkungan pengujian, bukan menyatakan keunggulan absolut terhadap konfigurasi penelitian lain.

4.10. Analisis Pengujian Real-Time

Pengujian dilakukan pada variasi pose depan, kiri, kanan, atas, dan bawah; jarak dekat dan sedang; pencahayaan cukup dan rendah; serta penggunaan aksesoris. Pada pose frontal dan pencahayaan cukup, wajah lebih mudah terdeteksi dan menghasilkan *similarity* yang stabil. Ketika wajah menoleh atau sebagian fitur tertutup, nilai *similarity* dapat menurun sehingga *temporal smoothing* diperlukan untuk mempertahankan label berdasarkan beberapa frame.

Pada deteksi multiwajah, setiap *bounding box* memiliki track ID dan riwayat prediksi sendiri. Informasi yang ditampilkan meliputi nama identitas, nilai cosine similarity, similarity margin, status Unknown atau Analyzing, jumlah wajah, jumlah track aktif, serta FPS. Pengujian menunjukkan sistem mampu memproses lebih dari satu wajah melalui kamera internal maupun eksternal, tetapi jarak yang terlalu jauh menyebabkan area wajah menjadi kecil dan mengurangi detail yang tersedia bagi FaceNet.

Kamera eksternal memberikan fleksibilitas penempatan, sedangkan kamera internal lebih sederhana untuk demonstrasi. Pada implementasi jaringan CCTV, tantangan tambahan akan mencakup protokol streaming, latensi jaringan, kompresi video, variasi sudut pemasangan, serta kualitas frame. Aspek tersebut berada di luar batas penelitian dan perlu diuji tersendiri sebelum penggunaan operasional.

Tabel 7. Analisis Skenario Pengujian Real-Time
Table 7. Analysis of Real-Time Test Scenarios

Skenario	Respons Sistem	Catatan
Wajah frontal	Deteksi dan similarity stabil	Kondisi paling representatif
Menoleh kiri/kanan	Tetap terdeteksi; similarity dapat menurun	Dipengaruhi sudut dan data enrollment
Pencahayaan rendah	CLAHE membantu, tetapi detail dapat berkurang	Memerlukan cahaya minimum yang memadai
Multiwajah	Setiap wajah diproses berdasarkan	Beban komputasi meningkat

	track	
Wajah tidak terdaftar	Diberi label Unknown	Bergantung threshold dan margin
Jarak terlalu jauh	Bounding box kecil dan prediksi kurang stabil	Resolusi kamera menjadi faktor penting

4.11. Analisis Black-Box dan Beta Testing

Black-box testing menunjukkan tujuh fungsi utama dapat berjalan sesuai keluaran yang diharapkan, yaitu pembacaan input kamera, deteksi satu wajah, deteksi beberapa wajah, pengenalan identitas terdaftar, pemberian label Unknown, penanganan variasi posisi dan pencahayaan, serta pemrosesan real-time. Pengujian ini memastikan integrasi antarmodul berjalan dari sisi pengguna tanpa menilai struktur internal kode.

Beta testing dilakukan kepada 10 pegawai. Rata-rata penilaian kemudahan penggunaan adalah 4,3; kecepatan sistem 4,0; akurasi pengenalan 4,2; tampilan antarmuka 4,5; dan stabilitas sistem 4,8. Respons didominasi skala setuju dan sangat setuju, sehingga prototipe dinilai mudah dipahami dan bermanfaat sebagai alat bantu identifikasi.

Hasil beta testing perlu dibaca sebagai evaluasi penerimaan awal, bukan sebagai bukti generalisasi kepada seluruh pengguna organisasi. Jumlah responden terbatas dan pengujian dilakukan pada prototipe. Meskipun demikian, hasil tersebut memberikan masukan bahwa informasi pada antarmuka dapat dipahami dan sistem dapat digunakan tanpa prosedur yang rumit. Penilaian kecepatan yang lebih rendah daripada aspek lainnya juga selaras dengan kebutuhan optimasi ketika sistem memproses beberapa wajah atau menggunakan perangkat dengan kemampuan terbatas.

4.12. Implikasi, Kelebihan, dan Keterbatasan

Kelebihan utama sistem adalah kemampuan memisahkan deteksi, ekstraksi fitur, dan database identitas. Arsitektur ini memudahkan penambahan identitas tanpa melatih ulang seluruh model. Mekanisme Unknown mengurangi risiko setiap wajah dipaksa masuk ke salah satu kelas terdaftar. Tracking dan smoothing juga meningkatkan keterbacaan hasil karena label tidak berubah hanya akibat fluktuasi satu frame.

Keterbatasan penelitian meliputi jumlah identitas dan data yang masih terbatas, belum adanya face alignment berbasis landmark, tracking yang hanya menggunakan IoU, dan belum diterapkannya face anti-spoofing. Sistem juga masih dipengaruhi pencahayaan sangat rendah, blur, oklusi, pose ekstrem, jarak terlalu jauh, serta resolusi kamera. Selain itu, label Unknown tidak dapat membedakan apakah wajah asing merupakan pengunjung yang sah, pegawai baru, atau upaya penyamaran.

Dalam konteks organisasi, penggunaan teknologi biometrik memerlukan kebijakan pengelolaan data, pembatasan akses, retensi data, pencatatan penggunaan, dan persetujuan yang sesuai. Penelitian ini hanya membahas aspek teknis prototipe dan tidak mencakup rancangan tata kelola biometrik secara menyeluruh. Oleh karena itu, implementasi operasional harus didahului kajian

keamanan informasi, privasi, kebutuhan pengguna, dan prosedur verifikasi manual.

4.13. Alur Operasional Sistem

Alur operasional dimulai ketika pengguna menentukan sumber kamera dan menjalankan aplikasi. OpenCV membuka stream video, kemudian setiap frame dibaca dan, apabila diperlukan, dibalik secara horizontal untuk menyesuaikan tampilan kamera internal. Frame diteruskan ke YOLOv8n-Face untuk memperoleh koordinat wajah. Bounding box yang lolos ambang kepercayaan kemudian disesuaikan menjadi persegi, ditambahkan margin, dan dipotong dari frame asli. Setiap crop menjalani *resize*, CLAHE, *sharpening*, konversi RGB, dan standardisasi sebelum masuk ke FaceNet.

Embedding hasil FaceNet dinormalisasi L2 dan dibandingkan dengan seluruh *centroid*. Sistem menyimpan nilai kandidat pertama, kandidat kedua, *best similarity*, dan *similarity margin*. Keputusan awal selanjutnya dimasukkan ke riwayat track. Apabila jumlah *vote* belum memenuhi batas minimum, antarmuka menampilkan *Analyzing*. Setelah bukti *temporal* mencukupi, label stabil ditampilkan bersama nilai *similarity*. Mekanisme tersebut membuat sistem tidak langsung mengunci identitas hanya berdasarkan satu frame yang mungkin *blur* atau tertutup sebagian.

Pada akhir pemrosesan frame, sistem menggambar *bounding box*, nama, status, *track ID*, dan informasi performa. Siklus ini diulang sampai pengguna menekan perintah keluar atau stream kamera berakhir. Dengan susunan modular, fungsi deteksi, *preprocessing*, *embedding*, pengenalan, *tracking*, dan visualisasi dapat diuji secara terpisah. Modularitas juga memudahkan penggantian kamera atau penyesuaian *threshold* tanpa mengubah keseluruhan *pipeline*.

Proses *enrollment* dijalankan terpisah dari inferensi *real-time*. Pengguna terlebih dahulu mengumpulkan citra pada folder berdasarkan identitas. Seluruh citra diproses menggunakan konfigurasi *preprocessing* yang sama dengan inferensi. *Embedding* kemudian disimpan dan diringkas menjadi *centroid*. Kesamaan konfigurasi antara *enrollment* dan *real-time* penting untuk menghindari perbedaan distribusi input yang dapat menurunkan *similarity*.

Kalibrasi dilakukan setelah *database centroid* terbentuk. Data validasi *known* dan *unknown* dievaluasi pada beberapa kombinasi *threshold* dan *margin*. Hasil setiap kombinasi dibandingkan berdasarkan kemampuan menerima *known* yang benar dan menolak *unknown*. *Parameter* terpilih disimpan dalam konfigurasi dan dimuat kembali pada aplikasi *real-time*. Dengan cara ini, keputusan penerimaan identitas mempunyai dasar empiris dari data validasi.

4.14. Perbandingan Hasil dengan Penelitian Terdahulu

Accuracy penelitian ini sebesar 89,12% lebih rendah daripada akurasi 93,8% pada Setiawan dkk. (2025) dan 93,7% pada Terampe & Arif Pramudwiatmoko (2025).

Perbedaan tersebut tidak dapat dipahami hanya sebagai peringkat performa karena setiap penelitian menggunakan jumlah identitas, pembagian data, perangkat, model pengenalan, dan skenario pengujian yang berbeda. Penelitian ini juga memasukkan kelas *Unknown* dan menerapkan aturan penolakan, sehingga sebagian wajah *known* yang meragukan sengaja ditolak daripada dipaksakan ke kelas terdaftar.

Dari sisi keamanan keputusan *open-set*, UDR 100% dan FAR 0% menunjukkan seluruh wajah asing pada data uji berhasil ditolak. Metrik ini tidak selalu dilaporkan pada penelitian yang berfokus pada klasifikasi *closed-set*. Dengan demikian, kontribusi utama penelitian ini bukan hanya nilai *accuracy*, tetapi juga mekanisme pembatasan keputusan melalui *threshold* dan *margin*. Konfigurasi tersebut menghasilkan *trade-off* berupa FRR 12,20%.

Kecepatan 12,45 FPS menunjukkan sistem lebih cepat daripada 7,2 FPS yang dilaporkan Terampe & Arif Pramudwiatmoko (2025), tetapi lebih lambat daripada waktu inferensi detektor 12,5 milidetik per frame pada Setiawan dkk. (2025). Angka 12,5 milidetik pada penelitian tersebut merujuk pada komponen deteksi, sedangkan 80,30 milidetik pada penelitian ini mencakup *pipeline* pengenalan secara keseluruhan. Oleh karena itu, unit dan cakupan pengukuran perlu diperhatikan sebelum melakukan perbandingan langsung.

Secara umum, hasil penelitian memperlihatkan bahwa detektor ringan, model *embedding pretrained*, dan pencocokan *centroid* dapat menghasilkan *prototipe* yang praktis untuk lingkungan kantor. Fokus pengembangan berikutnya perlu diarahkan pada perluasan populasi, pengujian lintas kamera, peningkatan *alignment*, dan evaluasi keamanan terhadap spoofing agar sistem lebih mendekati kebutuhan operasional.

5. KESIMPULAN

Penelitian berhasil membangun sistem *real-time face recognition* berbasis YOLOv8n-Face dan FaceNet InceptionResnetV1 *pretrained* VGGFace2. Citra wajah diproses melalui *square cropping*, *margin*, *resize*, CLAHE, *sharpening*, konversi warna, standardisasi, dan normalisasi L2. Identitas ditentukan menggunakan *cosine similarity* terhadap *centroid* dengan *similarity threshold* dan *similarity margin*, sedangkan wajah yang tidak memenuhi kriteria diberi label *Unknown*.

Pengujian menghasilkan *accuracy* 89,12%, KRR 87,79%, UDR 100%, FAR 0%, FRR 12,20%, dan kecepatan sekitar 12,45 FPS. Sistem mampu memproses beberapa wajah secara bersamaan, menolak wajah tidak terdaftar, dan menstabilkan prediksi melalui *tracking IoU* dan *temporal smoothing*. Dengan demikian, sistem dapat digunakan sebagai prototipe pendukung identifikasi serta monitoring keberadaan individu di lingkungan kerja.

6. SARAN

Pengembangan selanjutnya perlu memperluas dataset berdasarkan individu, sesi, pose, pencahayaan, jarak, latar belakang, oklusi, dan perangkat kamera. Data *Unknown*

untuk kalibrasi dan pengujian harus dipisahkan. Face alignment berbasis landmark dapat ditambahkan, sedangkan IoU tracking dapat ditingkatkan menggunakan ByteTrack atau DeepSORT. Sistem juga perlu dilengkapi liveness detection, keamanan penyimpanan data biometrik, dashboard monitoring, pencatatan waktu dan lokasi kamera, serta integrasi kamera jaringan.

7. REFERENSI

- Nurlita, B. W., Winarno, S., Nugraha, A., Najiib, A., Muttaqin, I., Zarifa, Y., Salsabila, P. M., & Fairuz Mumtaz, G. (2024). *COMPARISON OF ARCFACE AND DLIB PERFORMANCE IN FACE RECOGNITION WITH DETECTION USING YOLOV8 PERBANDINGAN KINERJA ARCFACE DAN DLIB DALAM PENGENALAN WAJAH DENGAN DETEKSI MENGGUNAKAN YOLOV8*. 9(2), 2024.
- Rachmansyah, R., Al, E., Rachmansyah, M. R., Rahmat, B., Wahanani, H. E., Pembangunan, U., Veteran, N., & Timur, J. (2025). *Implementation of FaceNet for Facial Search Function in Reverse Image Search System-Muhammad Implementation of FaceNet for Facial Search Function in Reverse Image Search System*. 07(04). <https://doi.org/10.54209/jatilima.v7i04.1880>
- Rianti, A., Wachid, N., Majid, A., Fauzi, A., Sistem, P., Informasi, T., Telekomunikasi, S., Upi Di Purwakarta, K., & Edu, A. (2023). *CRISP-DM: Metodologi Proyek Data Science*.
- Setiawan, A. T., Prihatmanto, A. S., Erlangga, E., Ariani, F., & Permana, A. (2025). Face Recognition dengan YOLO 8 dan Convolutional Neural Network (CNN). *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, 15(2), 213. <https://doi.org/10.36448/expert.v15i2.4658>
- Terampe, G. C., & Arif Pramudwiatmoko. (2025). FACIAL RECOGNITION PERFORMANCE EVALUATION WITH YOLOV8, ARCFACE, AND SVM IN A CONTACTLESS EMPLOYEE ATTENDANCE SYSTEM. *Jurnal Riset Informatika*, 8(1), 108–120. <https://doi.org/10.34288/jri.v8i1.465>
- Terven, J., & Cordova-Esparza, D. (2024). *A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS*. <https://doi.org/10.3390/make5040083>
- Wang, M., & Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, 429, 215–244. <https://doi.org/https://doi.org/10.1016/j.neucom.2020.10.081>