

ANALISIS PERFORMA *RESNET-50* UNTUK *PERSON RE-IDENTIFICATION* PADA LINGKUNGAN LINTAS KAMERA

^{1*}Muhammad Abdul Khafid, ²Wahyu Cahyo Utomo, ³Danang Wahyu Widodo

¹ Teknik Informatika, Universitas Nusantara PGRI Kediri
² Teknik Informatika, Universitas Nusantara PGRI Kediri
³ Teknik Informatika, Universitas Nusantara PGRI Kediri
E-mail: ¹_1abdu18khafid@gmail.com, ²wahyu.utomo@unpkdr.ac.id,
³danangwahyuwidodo@unpkediri.ac.id

Penulis Korespondens : Muhammad Abdul Khafid

Abstrak—Masalah utama pemantauan keamanan kampus lintas kamera adalah hilangnya jejak identitas visual pejalan kaki akibat topologi jaringan *non-overlapping* dan tingginya variasi intra-kelas. Untuk mengatasi masalah ini, penelitian mengevaluasi metode *Person Re-Identification (Re-ID)* menggunakan arsitektur *ResNet-50* standar. Proses pencocokan identitas dilakukan melalui ekstraksi fitur visual dan perhitungan *Cosine Similarity* antara citra kueri dan galeri. Hasil pengujian pada 20 target di lingkungan kampus menunjukkan performa pencarian awal yang sangat baik, dengan *Rank-1 Accuracy* mencapai 100% pada 19 target dan metrik *Recall* 0,9466. Namun, sistem mencatat *Precision* yang sangat rendah (0,09-0,13) dan *mAP* tertinggi 0,3979. Kesimpulannya, model *ResNet-50* standar andal menempatkan target utama di peringkat pertama, tetapi masih menghasilkan banyak *false positive* karena gagal membedakan fitur pejalan kaki berpakaian mirip di bawah variasi pencahayaan ekstrem.

Kata Kunci— *Person Re-Identification*, Analisis Performa, *ResNet-50*, *Cosine Similarity*, Lintas Kamera.

Abstract—The main problem in cross-camera campus security monitoring is the loss of pedestrian visual identity traces due to non-overlapping network topologies and high intra-class variations. To address this, this study evaluates the *Person Re-Identification (Re-ID)* method using a standard *ResNet-50* architecture. Identity matching was performed through visual feature extraction and *Cosine Similarity* calculations between query and gallery images. Testing on 20 targets in a campus environment demonstrated excellent initial retrieval, with *Rank-1 Accuracy* reaching 100% on 19 targets and a *Recall* of 0.9466. However, the system recorded very low *Precision* (0.09-0.13) and a maximum *mAP* of 0.3979. In conclusion, standard *ResNet-50* is reliable in ranking the main target first but generates numerous false positives due to its failure to distinguish features of similarly dressed pedestrians under extreme lighting variations.

Keywords— *Person Re-Identification*, Performance Analysis, *ResNet-50*, *Cosine Similarity*, Cross-Camera.

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Sistem keamanan dan pemantauan berbasis *Closed Circuit Television (CCTV)* telah menjadi infrastruktur standar di berbagai fasilitas publik, institusi pendidikan, dan pusat

perbelanjaan. Di lingkungan institusi pendidikan seperti Universitas Nusantara PGRI Kediri, kamera pemantau dipasang di berbagai titik strategis, seperti gerbang masuk, area parkir, dan lorong gedung perkuliahan. Sistem ini bertujuan untuk memantau pergerakan sivitas akademika, menjaga ketertiban, serta menyediakan rekaman forensik ketika terjadi insiden atau pelanggaran. Meskipun cakupan kamera semakin luas, topologi jaringan *CCTV* pada umumnya masih bersifat *non-overlapping*, yaitu bidang pandang antar kamera tidak saling bertumpang tindih dan dipisahkan oleh jarak maupun waktu tempuh tertentu[1].

Kondisi tersebut menimbulkan tantangan besar dalam manajemen keamanan karena identitas visual seseorang akan terputus ketika keluar dari jangkauan satu kamera dan masuk ke kamera lainnya. Hingga saat ini, proses pencocokan dan pelacakan antar kamera masih banyak dilakukan secara manual[2]. Petugas keamanan harus memutar ulang rekaman dari berbagai kamera, mencocokkan stempel waktu (*timestamp*), lalu mengandalkan pengamatan visual untuk memastikan apakah individu pada kamera yang berbeda merupakan orang yang sama. Proses manual ini tidak hanya menguras waktu dan konsentrasi, tetapi juga kurang efisien serta rentan terhadap human error, terutama ketika individu yang dilacak menggunakan atribut umum seperti jas almamater atau pakaian berwarna gelap[3].

Perkembangan pesat di bidang *Computer Vision* dan *Deep Learning* membuka peluang untuk mengotomatisasi proses tersebut melalui teknologi *Person Re-Identification (Re-ID)*. *Person Re-ID* merupakan metode komputasi yang dirancang untuk mengidentifikasi individu yang sama dari sekumpulan citra yang diambil oleh kamera berbeda dengan posisi, sudut pandang, dan kondisi pencahayaan yang beragam[4]. Fokus utama *Re-ID* bukan pada pengenalan wajah secara spesifik karena wajah sering kali tidak terlihat jelas akibat keterbatasan resolusi *CCTV* atau jarak pengambilan gambar yang jauh. Sebaliknya, metode ini menitikberatkan pada pengenalan fitur penampilan secara keseluruhan (*appearance-based features*), seperti warna pakaian, bentuk tubuh, dan barang bawaan[5].

Namun, implementasi *Re-ID* di lingkungan nyata menghadapi tantangan berupa variasi intra-kelas yang tinggi. Individu yang sama dapat terlihat sangat berbeda pada dua kamera akibat perubahan pencahayaan, perbedaan sudut pandang, oklusi, maupun perubahan postur tubuh saat berjalan[6]. Di sisi lain, sistem juga menghadapi masalah kesamaan inter-kelas, yaitu ketika dua individu berbeda memiliki penampilan yang sangat mirip karena menggunakan pakaian dengan warna atau model yang serupa[7]. Untuk mengatasi permasalahan tersebut, penelitian terdahulu di bidang *deep learning* umumnya berfokus pada pengembangan arsitektur jaringan yang kompleks, penambahan *attention mechanism*, maupun normalisasi spesifik domain guna meningkatkan akurasi pada *dataset* publik seperti *Market-1501* dan *DukeMTMC-reID*[8],[9]. Akan tetapi, penerapan model kompleks pada sistem keamanan kampus sering terkendala oleh keterbatasan infrastruktur komputasi.

[10][11] Oleh karena itu, penelitian ini menggunakan pendekatan yang lebih praktis dengan memfokuskan analisis pada kemampuan representasi fitur menggunakan arsitektur *Convolutional Neural Network (CNN) ResNet-50* standar. Penggunaan *ResNet-50* standar dipilih untuk mengevaluasi sejauh mana kemampuan dasar fitur residual dalam menangani variasi pencahayaan dan pose pada lingkungan kampus tanpa tambahan modul komputasi khusus[10]. Selain itu, penggunaan *dataset* citra pejalan kaki yang telah dipotong (*pre-cropped*) membuat penelitian ini berfokus secara khusus pada performa ekstraksi dan pencocokan fitur visual[11]. Dengan demikian, penelitian ini memberikan gambaran objektif mengenai batas kemampuan arsitektur dasar dalam skenario *Re-ID* nyata tanpa rekayasa *dataset* yang kompleks.

Gap penelitian yang ingin diisi adalah masih terbatasnya literatur yang menganalisis secara kritis keberhasilan maupun kegagalan model *CNN* dasar, seperti *ResNet-50* standar, ketika diterapkan pada data riil lintas kamera di lingkungan kampus tanpa rekayasa *dataset* khusus. Tujuan utama penelitian ini adalah melakukan analisis mendalam terhadap performa implementasi metode *Re-ID* menggunakan *ResNet-50* standar dengan menggunakan metrik evaluasi seperti *Accuracy*, *Precision*, *Recall*, *F1-Score*, *Rank-1 Accuracy*, dan *mean Average Precision (mAP)*. Melalui analisis empiris tersebut, kelemahan arsitektur standar dalam membedakan target dan *false positive* dapat dipetakan secara lebih jelas sehingga hasil penelitian ini diharapkan dapat menjadi rekomendasi ilmiah dalam pengembangan sistem pengenalan identitas pejalan kaki yang efektif, terjangkau, dan rasional untuk diterapkan pada lingkungan kampus. Dengan demikian, riset ini mengisi celah pengetahuan antara pengembangan model *Re-ID* yang semakin kompleks dengan kebutuhan praktis akan solusi yang sederhana dan mudah direproduksi.

II. METODE

A. Alur Evaluasi *Person Re-Identification*

Penelitian ini dirancang untuk menganalisis performa ekstraksi fitur visual pejalan kaki menggunakan arsitektur *ResNet-50* standar. Alur kerja sistem secara keseluruhan mencakup pengumpulan *dataset* citra pejalan kaki yang telah dipotong (*pre-cropped*), ekstraksi fitur menggunakan *ResNet-50* standar untuk menghasilkan representasi vektor fitur, dan kalkulasi jarak kemiripan menggunakan metode *Cosine Similarity* antara citra kueri dan citra galeri. Pendekatan ini dipilih untuk mengisolasi kontribusi arsitektur residual murni tanpa pengaruh modul tambahan seperti *attention mechanism* atau mekanisme *re-ranking* yang sering digunakan pada sistem *Re-ID* kontemporer. Setiap tahapan dalam alur kerja ini dirancang untuk meminimalkan variabel pengganggu sehingga hasil analisis performa benar-benar mencerminkan kemampuan intrinsik dari arsitektur yang diuji.

B. Skenario dan *Dataset* Pengujian

Data untuk penelitian ini dikumpulkan langsung dari lingkungan Universitas Nusantara PGRI Kediri. Pengambilan sampel citra melibatkan dua unit kamera pengawas dengan lokasi dan karakteristik lingkungan yang berbeda secara signifikan (lintas kamera): Kamera 1 diposisikan di dalam gedung Fakultas Teknik dan Ilmu Komputer (FTIK), sedangkan Kamera 2 diposisikan di area terbuka luar gedung yang menghadap ke jalan utama dan area parkir kampus. Perbedaan karakteristik lingkungan antara kedua kamera ini memberikan variasi pencahayaan dan sudut pandang yang ekstrem, sehingga merepresentasikan tantangan nyata dalam skenario *Re-ID* lintas kamera. Pemilihan dua lokasi dengan karakteristik yang kontras ini bertujuan untuk menguji secara ketat kemampuan generalisasi model dalam kondisi pencahayaan dan latar belakang yang berbeda secara signifikan.

1. Kamera 1 (Kueri)

Kamera 1 ditempatkan di area dalam gedung yang berfungsi sebagai lokasi pengambilan citra target (kueri) di mana subjek pertama kali teramati. Kondisi pencahayaan dipengaruhi oleh pantulan cahaya dalam gedung atau pencahayaan buatan, dengan sudut kamera sedikit lebih menukik. Lingkungan perekaman ini mencakup lorong dan area sekitar ruang kuliah sehingga subjek tampak dalam berbagai pose dan arah gerakan. Variasi iluminasi dan sudut pandang yang dihasilkan menjadi tantangan tersendiri bagi sistem dalam mengekstraksi fitur visual yang konsisten.

2. Kamera 2 (Galeri)

Bertindak sebagai database pencarian (galeri). Di area ini, terekam ratusan citra pejalan kaki yang lewat untuk dijadikan basis data pembanding (*distractors*). Kondisi pencahayaan lebih natural dan terang karena terpapar sinar matahari luar ruangan (*outdoor*). Perbedaan kondisi pencahayaan antara Kamera 1 dan Kamera 2 menjadi tantangan utama yang menguji ketangguhan representasi fitur arsitektur *ResNet-50* standar dalam skenario lintas kamera.

Evaluasi dilakukan terhadap 20 individu target berbeda yang diklasifikasikan sebagai citra kueri (target_1 hingga target_20). Setiap target dipotong (cropped) secara manual dari rekaman kamera untuk memastikan fokus evaluasi benar-benar pada kemampuan pencocokan fitur visual, bukan pada kinerja deteksi objek. Citra-citra ini kemudian dicocokkan dengan galeri gambar pada Kamera 2 untuk mengukur kemampuan identifikasi ulang sistem. Proses pencocokan dilakukan dengan menghitung skor *Cosine Similarity* antara vektor fitur setiap kueri terhadap seluruh vektor fitur citra dalam galeri, kemudian hasilnya diurutkan dari skor tertinggi ke terendah.

C. Ekstraksi Fitur Menggunakan *ResNet-50*

Citra pejalan kaki pada *dataset* diubah ukurannya (resize) agar memiliki dimensi matriks yang seragam, yaitu 256 X 128 piksel, sebelum diumpankan ke model pengekstraksi fitur. Ukuran ini dipilih karena merupakan standar umum pada tugas *Re-ID* berbasis *CNN* yang menyeimbangkan antara resolusi informasi visual dan efisiensi komputasi. Penelitian ini menggunakan arsitektur *ResNet-50* (*Residual Network 50-layers*). *ResNet-50* memiliki 50 lapisan konvolusional yang dilengkapi dengan mekanisme *skip connection* untuk mengatasi masalah *vanishing gradient* pada jaringan yang dalam.

ResNet-50 dirancang dengan blok-blok konvolusi yang memiliki koneksi Lewatan (*skip connection*). Konsep ini memungkinkan jaringan untuk mempertahankan informasi gradien dari lapisan awal hingga akhir tanpa mengalami masalah degradasi gradien (*vanishing gradient*). Proses konvolusi pada jaringan ini mengekstraksi peta fitur spasial yang menangkap pola warna, tekstur, dan bentuk pakaian pejalan kaki. Pada tahap akhir arsitektur, sebuah operasi *Global Average Pooling* (GAP) diterapkan untuk memadatkan peta fitur multi-dimensi menjadi representasi vektor satu dimensi (*feature vector*) dengan dimensi panjang sebesar 2048 nilai. Vektor ini berfungsi sebagai deskriptor visual utama dari setiap citra pejalan kaki.

D. Penghitungan *Cosine Similarity*

Untuk menentukan tingkat keserupaan identitas antara citra kueri dari Kamera 1 dengan citra galeri di Kamera 2, sistem membandingkan vektor fitur keduanya menggunakan metode *Cosine Similarity*. *Cosine Similarity* merupakan metrik keserupaan yang banyak digunakan dalam *information retrieval* karena kemampuannya mengukur orientasi relatif antar vektor tanpa terpengaruh oleh magnitudo masing-masing vektor. Metode ini menghitung nilai kosinus dari sudut yang terbentuk antara dua buah vektor fitur dalam ruang berdimensi 2048. Vektor fitur berdimensi 2048 ini diperoleh dari layer *global average pooling* pada arsitektur *ResNet-50* yang merepresentasikan fitur visual tingkat tinggi dari citra pejalan kaki.

Misalkan vektor fitur dari citra kueri adalah A dan vektor fitur citra galeri adalah B, rumus kalkulasi keserupaan kosinus dinyatakan sebagai berikut:

$$COSINE\ SIMILARITY(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

Hasil perhitungan ini menghasilkan skor desimal dalam rentang -1 hingga 1. Skor yang mendekati nilai 1 mengindikasikan bahwa kedua vektor memiliki arah yang sangat mirip, yang mengindikasikan tingkat kemiripan identitas visual yang tinggi antar-citra. Berdasarkan nilai skor tersebut, seluruh citra dalam database galeri diurutkan (*sorting*) dari nilai tertinggi ke terendah untuk membentuk daftar peringkat (rank retrieval).

E. Metrik Evaluasi Performa

Analisis kinerja ekstraksi fitur visual dievaluasi menggunakan sejumlah metrik standar dalam temu balik informasi (*information retrieval*), yaitu:

1. *Accuracy*

Accuracy merupakan metrik yang digunakan untuk mengukur proporsi prediksi yang dilakukan sistem secara benar terhadap seluruh data pengujian. Nilai ini menunjukkan seberapa baik sistem dalam membedakan antara citra target dan non-target secara keseluruhan. Semakin tinggi nilai *Accuracy*, maka semakin tinggi pula tingkat ketepatan sistem dalam melakukan klasifikasi atau pencocokan identitas antar kamera. Namun, dalam konteks *Re-ID* dengan jumlah data negatif yang jauh lebih banyak daripada data positif, *Accuracy* saja tidak cukup untuk menggambarkan performa sistem secara utuh sehingga diperlukan metrik tambahan yang lebih spesifik.

2. *Precision*

Precision menunjukkan kemampuan sistem untuk menghindari kesalahan pelabelan sampel negatif (orang lain) sebagai target positif. Dengan kata lain, *Precision* mengukur seberapa akurat sistem ketika mengklaim suatu kecocokan identitas. Metrik ini penting untuk memastikan bahwa hasil identifikasi yang dinyatakan cocok benar-benar merupakan target yang dicari. Nilai *Precision* yang rendah mengindikasikan bahwa sistem sering salah dalam mengidentifikasi individu yang tidak terkait sebagai target yang sama, yang dapat menurunkan tingkat kepercayaan pengguna terhadap hasil identifikasi sistem.

$$PRECISION = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

3. *Recall*

Recall merupakan kemampuan sistem dalam menemukan kembali seluruh citra target yang benar di dalam basis data galeri. Metrik ini mengukur sensitivitas sistem terhadap keberadaan target di seluruh daftar peringkat, bukan hanya pada peringkat teratas. Nilai *Recall* yang tinggi menunjukkan bahwa sistem mampu menemukan sebagian besar identitas target yang tersedia tanpa banyak melewatkan data yang relevan. *Recall* yang tinggi sangat penting dalam aplikasi keamanan di mana kehilangan jejak target (*false negative*) dapat berakibat fatal pada proses pemantauan.

$$RECALL = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

4. *F1-Score*

F1-Score merupakan nilai rata-rata harmonik yang menyeimbangkan antara *Precision* dan *Recall*. Metrik ini digunakan untuk memberikan gambaran performa sistem secara lebih seimbang, terutama ketika terdapat ketidakseimbangan jumlah data positif dan negatif. Nilai *F1-Score* yang tinggi menunjukkan bahwa sistem tidak hanya mampu menemukan target dengan baik, tetapi juga mampu meminimalkan kesalahan identifikasi terhadap individu lain. Dengan demikian, *F1-Score* menjadi metrik yang paling representatif untuk mengevaluasi kinerja sistem *Re-ID* secara holistik dalam skenario dunia nyata.

5. *Rank-1 Accuracy*

Rank-1 Accuracy merupakan metrik yang mengukur apakah citra yang berada pada peringkat pertama (skor kemiripan tertinggi) adalah benar citra target yang dicari, bernilai 100% jika benar dan 0% jika salah. Metrik ini menjadi indikator paling intuitif dalam sistem *Re-ID* karena mencerminkan pengalaman langsung pengguna saat melihat hasil pencarian teratas. Nilai *Rank-1 Accuracy* yang tinggi menunjukkan bahwa sistem secara konsisten menempatkan target yang benar pada posisi teratas daftar peringkat. Metrik ini sangat relevan untuk skenario peringatan dini di mana operator hanya perlu melihat satu hasil teratas untuk mengambil keputusan.

6. *mean Average Precision (mAP)*

Mean Average Precision (mAP) menilai ketepatan posisi seluruh citra target yang benar di dalam daftar peringkat pencarian secara keseluruhan. Nilai *mAP* yang tinggi menunjukkan bahwa semua citra target yang benar terkonsentrasi di peringkat-peringkat teratas. Berbeda dengan *Rank-1 Accuracy* yang hanya melihat satu peringkat terbaik, *mAP* memberikan evaluasi yang lebih komprehensif terhadap kualitas seluruh daftar peringkat. Metrik ini sangat berguna untuk membandingkan performa sistem *Re-ID* secara lebih adil karena mempertimbangkan distribusi semua citra target dalam daftar pencarian.

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan data hasil eksperimen yang komprehensif, diikuti dengan analisis mendalam terkait performa sistem berdasarkan metrik-metrik evaluasi yang telah ditetapkan. Data disajikan dalam bentuk tabel rekap yang memuat nilai *Accuracy*, *Precision*, *Recall*, *F1-Score*, *Rank-1 Accuracy*, dan *mAP* untuk masing-masing dari 20 target pengujian. Pembahasan difokuskan pada pemahaman perilaku arsitektur *ResNet-50* standar saat menghadapi tantangan variasi lintas kamera. Analisis dilakukan secara bertahap, dimulai dari interpretasi setiap metrik secara individual hingga sintesis temuan secara keseluruhan untuk memberikan gambaran yang utuh tentang kekuatan dan kelemahan arsitektur standar ini.

A. Data Hasil Evaluasi

Sistem diuji dengan menyajikan 20 gambar target (*query*) secara bergantian. Sistem kemudian melakukan pencarian di dalam galeri Kamera 2 yang memuat ratusan gambar distractor pejalan kaki. Tabel 1 menampilkan rekapitulasi data hasil perhitungan metrik evaluasi untuk masing-masing target uji. Hasil evaluasi disusun secara berurutan berdasarkan nomor target untuk memudahkan analisis perbandingan antar individu.

Tabel 1 Hasil Evaluasi Performa Re-Id Berbasis Resnet-50

Target	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	F1 - Score	<i>mAP</i>	<i>Rank-1</i>
target_1.png	0,5673	0,1315	0,5922	0,2152	0,2755	100
target_10.jpg	0,4575	0,0945	0,5146	0,1596	0,2027	100
target_11.jpg	0,4302	0,1073	0,6408	0,1838	0,2894	100
target_12.jpg	0,2859	0,1041	0,8058	0,1843	0,1738	100
target_13.jpg	0,2411	0,096	0,7816	0,171	0,1894	100
target_14.jpg	0,2776	0,1083	0,8592	0,1924	0,3618	100
target_15.jpg	0,2329	0,1067	0,9029	0,1908	0,3979	100
target_16.jpg	0,4375	0,1356	0,8592	0,2343	0,2914	100
target_17.jpg	0,3267	0,1134	0,8398	0,1999	0,1777	100
target_18.jpg	0,299	0,1023	0,7718	0,1807	0,2886	100
target_19.jpg	0,2256	0,1039	0,8835	0,186	0,1927	100
target_2.png	0,4473	0,1124	0,6553	0,1919	0,209	0
target_20.jpg	0,3281	0,1236	0,9369	0,2183	0,1741	100
target_3.jpg	0,2265	0,1063	0,9078	0,1903	0,3287	100
target_4.png	0,3004	0,1183	0,9272	0,2098	0,2225	100
target_5.png	0,4045	0,1178	0,7621	0,204	0,3076	100
target_6.png	0,3549	0,1142	0,8058	0,2001	0,2543	100
target_7.jpg	0,3554	0,1291	0,9466	0,2273	0,279	100
target_8.jpg	0,4745	0,1333	0,7718	0,2273	0,2204	100
target_9.jpg	0,4614	0,1254	0,733	0,2142	0,3007	100

Sistem diuji dengan menyajikan 20 gambar target (*query*) secara bergantian. Sistem kemudian melakukan pencarian di dalam galeri Kamera 2 yang memuat ratusan gambar distractor pejalan kaki. Tabel 1 menampilkan rekapitulasi data hasil perhitungan metrik evaluasi untuk masing-masing target uji. Hasil evaluasi disusun secara berurutan berdasarkan nomor target untuk memudahkan analisis perbandingan antar individu.

B. Analisis Ketimpangan *Recall* dan *Precision*

Berdasarkan data pada Tabel 1, fenomena paling mencolok dari implementasi arsitektur standar ini adalah terjadinya *trade-off* (ketimpangan) yang sangat ekstrem antara metrik *Recall* dan *Precision*. Namun, sensitivitas pencarian yang tinggi ini diiringi dengan nilai *Precision* yang jatuh sangat rendah. Dari keseluruhan 20 kueri pengujian, nilai *Precision* hanya berada pada rentang 0,0945 hingga maksimal 0,1356. Nilai di bawah 0,20 menunjukkan bahwa dari 100% gambar yang diprediksi mirip dengan target oleh sistem, kurang dari 20% gambar yang benar-benar merupakan target, sedangkan lebih dari 80% gambar sisanya adalah orang lain (*False Positive*).

Nilai *Recall* sistem tergolong tinggi hingga sangat impresif. Titik tertinggi dicapai pada objek target_7.jpg dengan *Recall* 0,9466, diikuti target_20.jpg sebesar 0,9369, dan target_4.png sebesar 0,9272. Tingginya nilai *Recall* ini memberikan penegasan empiris bahwa *ResNet-50* memiliki kemampuan penelusuran identitas yang sangat peka. Sistem berhasil memanggil ulang atau menemukan hampir seluruh citra target yang sah yang tersembunyi di lautan data galeri Kamera 2. Model tidak kehilangan target utamanya.

Analisis teknis dari fenomena ini bersumber pada cara kerja *ResNet-50* itu sendiri. Arsitektur ini sangat kuat dalam mengekstraksi distribusi makro warna, misalnya area baju putih, area celana hitam. Di lingkungan kampus, terdapat puluhan mahasiswa yang memakai kombinasi pakaian umum seperti ini. Karena tidak dilatih dengan penalti yang mengkhususkan perbedaan mikroskopis (seperti tekstur kain, keberadaan tas ransel tipis, atau corak kecil), *ResNet-50* memukul rata semua pejalan kaki berkemeja putih ke dalam satu kluster fitur yang berdekatan. Akibatnya, sistem terlalu mudah menganggap orang lain sebagai target, yang menyebabkan lonjakan angka *False Positive* dan menggerus nilai *Precision* secara drastis.

Rendahnya *Precision* dan tingginya *Recall* ini secara matematis menarik nilai *F1-Score* sistem untuk tertahan pada angka yang relatif rendah, yaitu pada rata-rata 0,15 hingga 0,23. Angka ini mengonfirmasi bahwa arsitektur *ResNet-50* standar belum mampu menyeimbangkan ketepatan dan kelengkapan hasil pencarian secara simultan. Nilai *F1-Score* yang rendah menjadi indikator bahwa sistem masih memerlukan peningkatan, baik dari segi representasi fitur maupun strategi pencocokan. Temuan ini memperkuat argumen bahwa arsitektur dasar *CNN* residual memiliki keterbatasan inheren ketika diterapkan pada tugas *Re-ID* lintas kamera tanpa modifikasi arsitektur tambahan.

C. Analisis Superioritas *Rank-1 Accuracy*

Meskipun nilai *Precision* global sangat rendah, sistem menunjukkan keunggulan yang luar biasa dari sudut pandang pemeringkatan awal (*early ranking*). Hal ini mengindikasikan bahwa meskipun sistem menghasilkan banyak *false positive*, target yang benar hampir selalu muncul di posisi teratas. Dari 20 target yang diuji, sistem sukses menyapu bersih nilai *Rank-1 Accuracy* sebesar 100% pada 19 target pengujian. Hanya satu target (target_10) yang gagal menempati peringkat pertama, yang menunjukkan bahwa secara umum arsitektur *ResNet-50* sangat andal dalam mengenali kemiripan visual tingkat tinggi antara kueri dengan citra target yang benar di galeri.

Arti praktis dari nilai ini adalah ketika operator memasukkan gambar kueri target, gambar yang muncul tepat di urutan ke-1 (yang paling dipercaya oleh model karena memiliki skor *Cosine Similarity* paling tinggi) adalah benar-benar gambar individu target yang dicari. Pencapaian ini membuktikan bahwa algoritma ini sangat fungsional jika ditugaskan sekadar untuk melokalisasi keberadaan seseorang secara cepat pada detik-detik awal. Meskipun di

urutan peringkat ke-2, ke-3, hingga ke-10 dipenuhi oleh *False Positive* (orang-orang berkemeja serupa), sistem masih memiliki sisa presisi lokal untuk menobatkan gambar yang paling identik di puncak klasemen. Hal ini membuat waktu tanggap operator keamanan dapat dipangkas signifikan dalam mengenali target utama.

D. Analisis Kegagalan Target dan Faktor Lingkungan

Meskipun nilai *Rank-1 Accuracy* secara umum menunjukkan hasil yang sangat baik, terdapat satu anomali yang mengalami kegagalan total, yaitu pada pengujian menggunakan target_2.png. Pada kueri tersebut, nilai *Rank-1 Accuracy* mencapai 0%, yang menunjukkan bahwa gambar dengan peringkat pertama bukan merupakan gambar dari target_2, melainkan individu lain. Berdasarkan analisis pada kolom rata-rata skor di Tabel 1, target_2.png memiliki nilai rata-rata skor target sebesar 0,3330, sedangkan rata-rata skor non-target atau orang lain sebesar 0,3100. Selisih kemiripan yang sangat tipis tersebut menunjukkan bahwa sistem mengalami ambiguitas tinggi dalam membedakan individu asli dengan individu lain yang memiliki karakteristik visual serupa.

Analisis visual terhadap kasus anomali ini menunjukkan beberapa faktor lingkungan yang menyebabkan kegagalan sistem. Faktor pertama adalah variasi pencahayaan lintas kamera, di mana Kamera 1 sebagai kueri berada di koridor gedung dengan intensitas cahaya rendah, sedangkan Kamera 2 sebagai galeri berada di luar ruangan dengan pencahayaan matahari yang sangat terang. Kondisi ini menyebabkan fitur visual pakaian target mengalami perubahan signifikan sehingga sulit dikenali secara konsisten oleh model *ResNet-50* yang belum memiliki mekanisme normalisasi khusus untuk menangani perubahan domain pencahayaan ekstrem. Faktor kedua adalah perbedaan sudut pandang *viewpoint* dan terjadinya oklusi. Perubahan posisi tubuh saat berjalan, misalnya dari posisi menyamping pada Kamera 1 menjadi posisi membelakangi pada Kamera 2, mengubah representasi visual tubuh sehingga ekstraktor fitur kesulitan menghubungkan geometri objek yang sama. Akibatnya, nilai *Cosine Similarity* pada target melemah dan peringkatnya dapat disalip oleh citra individu lain yang memiliki kemiripan visual lebih tinggi.

E. Analisis Nilai *mean Average Precision (mAP)*

Metrik terakhir yang dianalisis adalah *mean Average Precision (mAP)*, yang mengevaluasi ketangguhan sistem dalam menyusun seluruh daftar peringkat pencarian. Metrik *mAP* memberikan gambaran yang lebih lengkap dibandingkan *Rank-1 Accuracy* karena mempertimbangkan posisi semua citra target yang benar di seluruh peringkat, bukan hanya posisi pertama. Berdasarkan pengujian, skor *mAP* berada di area nilai yang moderat ke bawah, dari batas terendah 0,1738 hingga tertinggi hanya 0,3979. Rentang nilai ini mengindikasikan bahwa meskipun sistem berhasil menempatkan target di peringkat pertama, sisa citra target yang sama cenderung tersebar di peringkat-peringkat bawah dan bercampur dengan banyak citra distractor.

Analisis dari temuan ini kembali menegaskan hipotesis bahwa *ResNet-50* kurang tangguh (less robust) sebagai model *Re-ID* komprehensif. Meskipun sistem berhasil meletakkan satu gambar yang benar di peringkat-1 (*Rank-1* 100%), ternyata gambar-gambar benar lainnya dari target yang sama terlempar jauh ke peringkat-peringkat bawah (misalnya di peringkat ke-40 atau ke-70). Peringkat-peringkat atas justru dibajak oleh citra pejalan kaki lain akibat minimnya penalti sistem terhadap fitur yang serupa tapi tak sama. Dengan demikian, penggunaan *ResNet-50* terbukti cukup baik untuk membuat sistem peringatan dini awal (menemukan jejak pertama objek), namun kurang mumpuni jika difungsikan untuk melacak

jejak objek tersebut di setiap detik rekaman (*tracking*) di dalam lautan galeri multi-kamera yang bervariasi.

IV. KESIMPULAN

Pemantauan keamanan kampus menghadapi tantangan besar berupa terputusnya jejak identitas visual pejalan kaki saat melintasi area dengan topologi kamera *non-overlapping*. Kondisi ini diperparah oleh tingginya tingkat variasi pencahayaan ekstrem lintas kamera dan kesamaan visual pakaian antar pejalan kaki (variasi intra-kelas), sehingga proses pelacakan manual menjadi sangat tidak efisien, menguras tenaga, dan rentan terhadap kesalahan pengamatan. Untuk mengatasi permasalahan tersebut, penelitian ini mengimplementasikan metode *Person Re-Identification (Re-ID)* dengan memanfaatkan arsitektur *Convolutional Neural Network (CNN) ResNet-50* standar. Model ini digunakan sebagai pengekstraksi fitur visual utama dari *dataset* citra pejalan kaki, yang kemudian tingkat kecocokannya dievaluasi secara komputasi menggunakan perhitungan jarak *Cosine Similarity* antara citra kueri dari satu kamera dengan sekumpulan citra galeri di kamera lainnya.

Berdasarkan hasil pengujian dan analisis mendalam, dapat disimpulkan bahwa implementasi arsitektur *ResNet-50* standar mampu beroperasi dengan sangat baik pada tahap penemuan kembali awal. Model ini memiliki keunggulan absolut pada metrik pencarian teratas, dibuktikan dengan keberhasilan mencapai 100% *Rank-1 Accuracy* pada 95% data uji (19 dari 20 target) serta nilai *Recall* yang sangat sensitif (rata-rata di atas 0,70). Namun demikian, analisis kritis menemukan bahwa pendekatan ini masih memiliki kekurangan substansial dalam membedakan fitur halus intra-kelas, yang ditunjukkan oleh jatuhnya metrik *Precision* ke level rendah (0,09 - 0,13) dan skor *mAP* moderat (tertinggi 0,3979). Kesenjangan ini merepresentasikan banyaknya *false positive* akibat ketidakmampuan model dasar menangkal pergeseran domain lingkungan nyata. Pengembangan penelitian di masa depan direkomendasikan untuk beralih mengintegrasikan lapisan khusus seperti *Instance-Batch Normalization (IBN)* atau modifikasi arsitektur berbasis *Transformer* agar diskriminasi fitur visual target menjadi jauh lebih kokoh.

DAFTAR PUSTAKA

- [1]Y.-J. Cho and K.-J. Yoon, "Distance-based camera network topology inference for person re-identification," *Pattern Recognit. Lett.*, vol. 125, pp. 220–227, 2019, doi: <https://doi.org/10.1016/j.patrec.2019.04.009>.
- [2]N. Narayan, N. Sankaran, S. Setlur, and V. Govindaraju, "Learning deep features for online person tracking using non-overlapping cameras: A survey," *Image Vis. Comput.*, vol. 89, pp. 222–235, 2019, doi: <https://doi.org/10.1016/j.imavis.2019.07.007>.
- [3]A. Zahra, N. Perwaiz, M. Shahzad, and M. M. Fraz, "Person Re-identification: A Retrospective on Domain Specific Open Challenges and Future Trends," 2022. [Online]. Available: <https://arxiv.org/abs/2202.13121>
- [4]N.-B. Nguyen, V.-H. Nguyen, T. D. Ngo, and K. M. T. T. Nguyen, "Person re-identification with mutual re-ranking," *Vietnam Journal of Computer Science*, vol. 4, no. 4, pp. 233–244, 2017, doi: 10.1007/s40595-016-0093-x.
- [5]C.-H. Chuang, T.-C. Huang, C.-W. Wang, J.-H. Lo, and C.-L. Lin, "Person Re-Identification Under Non-Overlapping Cameras Based on Advanced Contextual Embeddings," *Algorithms*, vol. 18, no. 11, 2025, doi: 10.3390/a18110714.
- [6]A. Zahra, N. Perwaiz, M. Shahzad, and M. M. Fraz, "Person Re-identification: A Retrospective on Domain Specific Open Challenges and Future Trends," 2022. [Online]. Available: <https://arxiv.org/abs/2202.13121>

- [7]A. Zahra, N. Perwaiz, M. Shahzad, and M. M. Fraz, “*Person Re-identification: A Retrospective on Domain Specific Open Challenges and Future Trends*,” 2022. [Online]. Available: <https://arxiv.org/abs/2202.13121>
- [8]C. H. Chuang, T. C. Huang, C. W. Wang, J. H. Lo, and C. L. Lin, “*Person Re-Identification Under Non-Overlapping Cameras Based on Advanced Contextual Embeddings*,” *Algorithms*, vol. 18, no. 11, Nov. 2025, doi: 10.3390/a18110714.
- [9]Y. Chen, Z. Fan, Z. Chen, and Y. Zhu, “*CA-Jaccard: Camera-aware Jaccard Distance for Person Re-identification*.”
- [10]A. Zahra, N. Perwaiz, M. Shahzad, and M. M. Fraz, “*Person Re-identification: A Retrospective on Domain Specific Open Challenges and Future Trends*,” 2022. [Online]. Available: <https://arxiv.org/abs/2202.13121>
- [11]Y.-J. Cho and K.-J. Yoon, “Distance-based camera network topology inference for *person re-identification*,” *Pattern Recognit. Lett.*, vol. 125, pp. 220–227, 2019, doi: <https://doi.org/10.1016/j.patrec.2019.04.009>.