

Analisis *Spatiotemporal* Pergerakan Pengunjung Ruang Tunggu Rumah Sakit Menggunakan YOLO dan *Multi-Object Tracking*

Ardiansyah^{1*}, Nisrina Akbar Rizky Putri¹, Rezyana Budi Syahputri²

¹ Fakultas Kesehatan dan Teknologi, S1 Teknologi Informasi, Universitas Muhammadiyah Klaten, Klaten, Indonesia

² Fakultas Kesehatan dan Teknologi, S1 Administrasi Kesehatan, Universitas Muhammadiyah Klaten, Klaten, Indonesia

INFORMASI ARTIKEL

Diterima Redaksi: 24 Agustus 2026
Revisi Akhir: 26 September 2026
Diterbitkan Online: 29 September 2026

KATA KUNCI

Data Driven
Deep Learning
Multi-Object Tracking
Spatiotemporal
YOLO

KORESPONDENSI (*)

E-mail: ardiansyah@umkla.ac.id

A B S T R A K

Peningkatan pemanfaatan layanan kesehatan sebesar 6,6% dibandingkan tahun sebelumnya meningkatkan tekanan terhadap kapasitas pelayanan, sementara standar waktu tunggu rawat jalan ditetapkan <60 menit dan pada kondisi aktual masih ditemukan waktu tunggu >130 menit. Kondisi tersebut menunjukkan kebutuhan akan informasi operasional yang objektif untuk memahami kepadatan dan pergerakan pengunjung secara real-time. Penelitian ini bertujuan mengembangkan analisis spatiotemporal pergerakan pengunjung ruang tunggu rumah sakit berbasis video Closed-Circuit Television (CCTV) menggunakan You Only Look Once (YOLO) dan Multi-Object Tracking (MOT), berbeda dengan pemanfaatan sensor dan pencatatan manual, keduanya rentan bias dalam mengekstraksi dinamika spatiotemporal seperti pola lintasan maupun pemetaan kepadatan area. Penelitian menggunakan YOLOv8n-seg dan YOLOv11n-seg yang dikombinasikan dengan Deep OC-SORT dan StrongSORT pada video dengan frame rate 10 FPS dan 15 FPS. Hasil evaluasi deteksi menunjukkan bahwa kedua model memiliki performa yang tinggi, dengan Precision 0,972 dan Recall 0,971 pada YOLOv8n-seg serta Precision 0,969 dan Recall 0,974 pada YOLOv11n-seg; mAP50 masing-masing mencapai 0,985 dan 0,986, sedangkan F1-Score keduanya sebesar 0,971. Hasil estimasi occupancy menunjukkan bahwa kombinasi YOLOv8n-seg dan Deep OC-SORT menghasilkan MAE 0,94 orang, RMSE 1,31 orang, MAPE 5,98%, dan R² 0,8218 pada 15 FPS. Hasil tracking selanjutnya digunakan untuk membentuk trajectory, heatmap, dan perubahan occupancy terhadap waktu serta diintegrasikan ke dalam dashboard analitik real-time. Penelitian ini menunjukkan bahwa integrasi YOLO dan MOT dapat mentransformasikan data CCTV menjadi informasi spasial dan temporal yang objektif untuk mendukung pemantauan kepadatan ruang tunggu rumah sakit.

PENDAHULUAN

Manajemen fasilitas kesehatan di zaman modern menekankan tentang efisiensi operasional dan manajemen waktu tunggu dari manajemen fasilitas kesehatan modern [1], [2]. Namun di Indonesia, kedua hal tersebut seringkali masih menjadi tantangan dasar pada Fasilitas Kesehatan [3]. Meskipun, Kementerian Kesehatan telah mengatur Standar Pelayanan Minimal Rumah Sakit pada Keputusan Menteri Kesehatan RI bernomor 129/Menkes/SK/II/2008 yaitu <60 menit untuk standar waktu tunggu rawat jalan [4].

Regulasi yang telah ditetapkan, memberikan beban pelayanan yang terus meningkat. Laporan monitoring bulanan Jaminan Kesehatan Nasional (JKN) yang diterbitkan 31 Juli 2025 mendapatkan adanya peningkatan pada pemanfaatan layanan kesehatan 6,6% jika dibandingkan dengan tahun sebelumnya. Peningkatan yang terjadi di fasilitas Kesehatan membentuk ketimpangan antara kapasitas layanan dan permintaan. Realita di lapangan masih ditemukan hasil waktu tunggu pasien >130 menit [5], [6] meskipun di wilayah perkotaan, sehingga berdampak pada kepuasan pengunjung

fasilitas kesehatan [3], [7]. Ketimpangan ini menuntut adanya manajemen operasional yang presisi dan berbasis data, bukan sekedar intuisi manajerial.

Lebih jauh, Ketersediaan data masih terbatas pada angka kunjungan, jumlah pasien di setiap poli, maupun laporan manual tanpa menggunakan data detail tentang pola spasial-temporal dari pengunjung [8]. Fasilitas Kesehatan membutuhkan pendekatan yang berbeda seperti pendekatan *data-driven* yang dihasilkan dari kondisi sesungguhnya [9], [10]. Pendekatan *data-driven* yang dihasilkan dari ekstraksi video menjadi pola kunjungan, pemetaan *bottleneck* [11], serta pengambilan keputusan berdasarkan bukti [8].

You Only Look Once (YOLO) merupakan algoritma deep learning [12], [13], [14] yang dapat digunakan untuk analisis spatiotemporal pengunjung menawarkan peluang strategis. Algoritma YOLO mampu mengekstraksi pola pergerakan [15],[16] durasi tunggu, serta dinamika kepadatan secara *real-time* dengan menggunakan data video seperti CCTV [17]. Sehingga dapat menghasilkan data fundamental untuk mendukung peningkatan efisiensi layanan [18], [19].

Penelitian [11] pada bidang kesehatan terkhusus pada alur pasien telah dilakukan namun penelitian [11] sekedar membandingkan data timestamps di komputer (EMR) dengan observasi manual. Selain itu ditemukan banyak data tidak akurat karena staf lupa input waktu secara *real-time*. Penelitian [20] serupa berhasil memantau kondisi kesehatan dan lokasi pasien secara *real-time* berbasis kamera, IoMT, dan YOLOv8. Namun, Ketergantungan terhadap pada alat bantu (*LED transmitter*) dapat membebani biaya pengadaan alat *wearable* dan kerumitan manajemen aset. Sehingga, penggunaan cctv menjadi solusi yang ditawarkan pada penelitian yang akan dilakukan.

Berdasarkan literatur yang telah dipaparkan, terlihat bahwa penelitian terdahulu masih bersifat parsial serta belum dalam kerangka analisis yang terpadu. Penelitian tentang *patient flow* dan efisiensi pelayanan lebih banyak mengandalkan pendekatan simulasi dan analisis data administrative [11]. Sementara itu, sistem cerdas menggunakan sensing/IoT untuk mendukung manajemen layanan telah dilakukan, namun memunculkan masalah dari sisi pembiayaan yang meningkat [20], [21].

Lebih lanjut, berdasarkan literatur menggambarkan bahwa sebagian penelitian terdahulu belum menghubungkan tugas seperti deteksi pengunjung [1], [22] pelacakan *multi-object* [23], analisis spatiotemporal dalam satu alur. Penelitian [24], [25] menggabungkan YOLO dan pelacakan mampu bekerja dengan baik. Begitu juga dengan penelitian [8] yang menambahkan heatmap dalam ruang lingkup pengawasan epidemiological. Oleh karena itu, penelitian ini menawarkan kebaruan melalui pengembangan pemantauan pengunjung menggunakan YOLO yang mengintegrasikan analisis spasial-temporal seperti Smart City [26] ke dalam lingkungan rumah sakit [27], guna menghasilkan data yang objektif, *real-time*, dan bebas dari intervensi manual.

Analisis spatiotemporal menuntut tingkat presisi spasial yang tinggi karena interaksi antarindividu serta objek yang berada di sekitarnya. Penggunaan *bounding box* memiliki kelemahan mendasar yaitu tumpang tindih (*overlap*) saat pengunjung duduk berdampingan sehingga dapat mendistorsi estimasi titik posisi sebenarnya [16], [24]. Oleh karena itu, pendekatan segmentasi instans menggunakan YOLOv8n-seg dan YOLOv11n-seg menjadi pendekatan yang lebih diunggulkan jika dibandingkan *bounding box* [16], penggunaan varian nano menjadi pilihan yang tepat meskipun segmentasi memiliki beban komputasi yang lebih tinggi. Sementara itu, Deep OC-SORT dan StrongSORT keduanya memiliki asosiasi serta pemanfaatan informasi identitas objek yang berbeda dalam mempertahankan identitas antar-frame, sehingga dapat digunakan untuk mengevaluasi pengaruh metode pelacakan terhadap kestabilan identitas objek, keberlanjutan lintasan serta menghitung objek di kondisi dinamis [23].

Dengan demikian, kebaruan penelitian ini terletak pada penerapan analisis spatiotemporal pergerakan pengunjung pada kondisi riil pada pelayanan kesehatan melalui integrasi YOLO dan *Multi-Object Tracking* (MOT). Penelitian ini mengolah hasil deteksi dan *tracking* menjadi informasi yang menggambarkan pergerakan individu melalui *trajectory*, distribusi aktivitas melalui *heatmap*, serta perubahan *occupancy* pengunjung sepanjang waktu berdasarkan video CCTV dengan frame rate 10 FPS dan 15 FPS. Selanjutnya, hasil analisis tersebut diintegrasikan ke dalam dashboard *real-time* yang menampilkan kondisi pengunjung, *ground truth*, hasil *tracking*, dan *error* sebagai informasi pendukung pemantauan kepadatan ruang tunggu.

TINJAUAN PUSTAKA

YOLO

You Only Look Once (YOLO) merupakan model *One-Stage-Detector* [28] yang seringkali digunakan karena YOLO memiliki kemampuan memberikan akurasi tinggi dengan kecepatan *inference real-time* serta kemudahan dalam pelatihan maupun implementasi [29], [30]. YOLO terus mengalami perkembangan yang menghasilkan berbagai versi diantaranya adalah YOLOv8 dan YOLOv11 dengan berbagai skala ukuran model [16]. YOLOv8 menggunakan pendekatan *anchor-free split* head serta rancangan *backbone* dan *neck* yang dioptimalkan dengan tujuan memberikan keseimbangan kecepatan inferensi dengan akurasi deteksi objek [16]. Sedangkan YOLOv11 berfokus pada peningkatan ekstraksi fitur dengan pembaharuan pada arsitektur model sebelumnya. Pembaharuan yang dilakukan yaitu modul C2PSA yang terintegrasi dengan mekanisme self-attention. Selain peningkatan ekstraksi fitur, YOLOv11 juga menggantikan modul C2F dengan C3K2 untuk efisiensi parameter yang lebih sedikit dari model sebelumnya [16].

Multi-Object Tracking

Multi-Object Tracking (MOT) merepresentasikan proses yang bertujuan melacak multiple objek secara berkesinambungan pada rangkaian frame dari video, tetapi tetap mempertahankan identitas objek yang sama disetiap frame. Pendekatan *tracking-by-detection*, *output* dari deteksi objek pada setiap frame video digunakan sebagai dasar untuk proses pelacakan sebuah objek. Selanjutnya objek yang telah terdeteksi pada frame di integrasikan melalui proses prediksi posisi objek, kesamaan objek, serta asosiasi data sehingga objek mendapatkan identitas objek yang konsisten pada rangkaian video [23], [24]. Pendekatan *tracking-by-detection* memungkinkan sistem mendapatkan informasi yang tidak hanya keberadaan dan posisi objek, tetapi keberlanjutan objek setiap *frame*.

Meskipun begitu, MOT sangat dipengaruhi oleh kemampuan model atau sistem mempertahankan identitas dari objek ketika objek mengalami perubahan posisi, okulasi, maupun objek yang tidak terdeteksi dalam waktu pendek. Algoritma Deep OC-SORT dan StrongSORT merupakan algoritma lanjutan dari algoritma pelacakan SORT [31], [32], pemanfaatan fitur *appearance* melalui *re-identification* untuk membandingkan karakteristik dari objek disetiap *frame* serta menjaga keberlanjutan identitas [23]. MOT dapat dipandang sebagai komponen yang mengubah hasil deteksi *per-frame* menjadi informasi pergerakan objek yang berkesinambungan, yang selanjutnya dapat digunakan untuk memperoleh *trajectory* dan menganalisis pola pergerakan objek dalam video [23], [24].

Analisis Spatiotemporal

Analisis *spatiotemporal* merupakan pendekatan yang menghubungkan antara informasi spasial dengan temporal. Kedua hal tersebut bertujuan untuk memahami pergerakan, keberadaan, dan perubahan dari sebuah objek per ruang dan waktu. Informasi Spasial menjelaskan karakteristik serta posisi objek pada suatu lokasi, sedangkan informasi temporal menjelaskan perubahan dan keberlanjutan objek disetiap frame atau sepanjang interval waktu [30], [33]. Keterkaitan antara spasial dan temporal dapat membentuk ataupun mempengaruhi pola obeservasi yang dihasilkan [34]. Analisis *spatiotemporal* dalam konteks analisis video dan *computer vision*, menghubungkan keduanya sehingga membentuk informasi posisi maupun pergerakan objek yang dianalisis setiap frame per *timestamp* [30], [33], [35].

Untuk memperjelas posisi serta kontribusi penelitian terhadap perkembangan objek deteksi, MOT, dan okupansi berbasis computer vision, terlihat pada Tabel 1.

Table 1. Penelitian Terdahulu

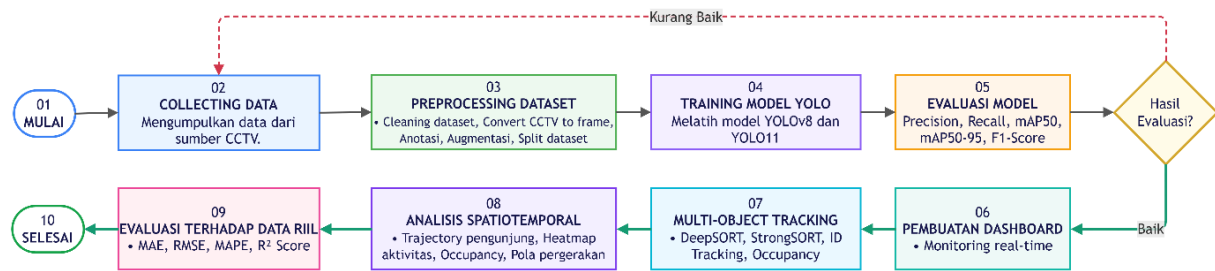
Judul Penelitian / Penulis	Metode	Objek Penelitian	Kekurangan	Posisi Penelitian Penulis
<i>Improving Patient Flow in a Healthcare Clinic Post COVID-19: A Data Validation and Exploratory</i>	Observational data, validasi timestamp data EMR, analisis statistic serta discrete-event simulation	UPC Primary care clinic	Ketergantungan pada data EMR menjadi kekurangan penelitian karena ditemukan banyak data tidak akurat karena staf lupa	Penelitian yang diusulkan memperkenalkan pengukuran otomatis waktu tunggu, rute, dan pola kepadatan dari CCTV. Sehingga

Judul Penelitian / Penulis	Metode	Objek Penelitian	Kekurangan	Posisi Penelitian Penulis
<i>Analysis Approach</i> [11]			input waktu secara <i>real-time</i> .	informasi yang didapatkan <i>real-time</i>
<i>Computer Vision Model for Analyzing Registry Workload CDC “Almazov National Medical Research Centre”</i> [22]	Model YOLOv8n task Object detection	Area registrasi CDC Almazov national medical research center	Fine-tuned YOLOv8 meningkatkan akurasi untuk deteksi pengunjung pada loket registrasi. Namun Penelitian sebelumnya tidak menggunakan MOT, tidak menghitung waktu tunggu, dan tidak melakukan analisis spatiotemporal dengan memanfaatkan <i>bounding box</i> .	Penelitian yang akan dilakukan mengisi gap tersebut dengan YOLO, MOT, heatmap. Penelitian yang dilakukan memperluas pendekatan dari deteksi dan menghitung objek, sedangkan untuk mempertahankan identitas objek menggunakan Deep OC-Sort dan StrongSORT. Sehingga analisis tidak hanya mencakup jumlah individu saja.
<i>Real-Time Human Group Detection and Clustering in Crowded Environments Using Enhanced Multi-Object Tracking</i> [24]	Integrasi YOLOv8 dan DeepSort	Individu dan kelompok pada Lingkungan bandara	YOLO dan MOT dapat digunakan untuk mendeteksi kelompok manusia dalam kerumunan, namun masih menggunakan <i>bounding box</i> sebagai representasi objek. Selain itu, peneliti mengakui bahwa occlusion dan ketidaklengkapan anotasi menjadi tantangan dalam penelitian tersebut.	Penelitian tersebut belum ada analisis waktu tunggu atau pemetaan ruangan; sehingga akan dilanjutkan ke analisis spatiotemporal operasional rumah sakit menggunakan <i>Instance Segmentation</i> .
<i>Integrating Object Detection and Advanced Analytics for Smart City Crowd Management</i> [26]	Integrasi YOLOv8 dan BYTETracker	Pejalan kaki pada Times Square, New York	Menggabungkan YOLO dengan analisis lanjutan seperti heatmap dan prediksi lintasan di lingkungan kota (Smart City).	Penelitian ini tidak memetakan pergerakan individu di fasilitas kesehatan; Penelitian yang akan dilakukan melakukan spatiotemporal analysis, dwell time, heatmap
<i>Enhancing Hospital Efficiency Through Web-Deployed Object Detection: A YOLOv8-Based Approach for Automating Healthcare Operations</i> [1]	Integrasi YOLOv8 dengan React.js dan TensorFlow.js	Operasional rumah sakit	YOLOv8 dapat mendeteksi berbagai objek rumah sakit untuk efisiensi operasional. Namun, hanya sekedar mendeteksi objek, tidak ada pelacakan lanjutan.	Penelitian ini tidak menganalisis pergerakan manusia atau alur pasien. Penelitian yang akan dilakukan mengintegrasikan deteksi manusia, tracking, spatiotemporal analytics sebagai fondasi data-driven

Judul Penelitian / Penulis	Metode	Objek Penelitian	Kekurangan	Posisi Penelitian Penulis
<i>Hospital crowdedness evaluation and in-hospital resource allocation based on image recognition technology</i> [10]	Multi-Fusion Convolutional Neural Network berbasis ResNet dan U-Net	Area pelayanan dan ruang tunggu Instalasi Gawat Darurat	Sampling video sangat rendah yaitu: 0.2 FPS sehingga sebagian besar individu dianggap tidak bergerak pada area tunggu. Sehingga Penelitian tersebut hanya menghitung jumlah orang, tidak melakukan tracking ID, tidak menghitung waiting time atau lintasan.	Penelitian yang diusulkan menambah MOT, heatmap, analisis waktu sehingga hasilnya dapat menjadi metrik manajerial. Penelitian ini memberikan informasi secara real-time dengan memanfaatkan instance segmentation serta multi <i>object tracking</i> .
<i>Enhanced DeepSORT and StrongSORT for Multicattle Tracking with Optimized Detection and Re-Identification</i> [23]	Tracking-by-detection menggunakan YOLOv9 dan DeepSORT serta StrongSORT.	Cattle/livestock	Penelitian tersebut masih menggunakan bounding box sebagai representasi objek serta mendapatkan penurunan performa ketika terjadi salah deteksi.	penelitian yang diusulkan mengaplikasikan MOT ini ke konteks layanan kesehatan dan mengubah output tracking menjadi indikator manajerial (<i>waiting time, bottleneck</i>). Pemanfaatan <i>instance segmentation</i> menjadi perbedaan penelitian yang dilakukan.
<i>Real-Time Patient Indoor Health Monitoring and Location Tracking with Optical Camera Communications on the Internet of Medical Things</i> [20]	<i>Optical Camera Communication</i> (OCC) berbasis YOLO untuk deteksi LED matriks yang merepresentasikan pasien. Data kesehatan dikirim melalui kamera ke edge server dan lokasi relatif pasien diperoleh dari beberapa kamera CCTV	Lingkungan indoor rumah sakit	berhasil memantau kondisi kesehatan dan lokasi pasien secara real-time berbasis kamera, IoMT, dan YOLOv8. Ketergantungan terhadap pada alat bantu (<i>LED transmitter</i>) dapat membebani biaya pengadaan alat wearable dan kerumitan manajemen aset.	Penelitian yang dilakukan memanfaatkan <i>visual appearance individu</i> secara langsung tanpa perangkat LED sebagai representasi objek. Model YOLO dan MOT digunakan untuk memperoleh spasial individu dan mempertahankan identitas setiap individu serta menganalisis okupansi maupun pergerakan secara spatiotemporal.

METODOLOGI

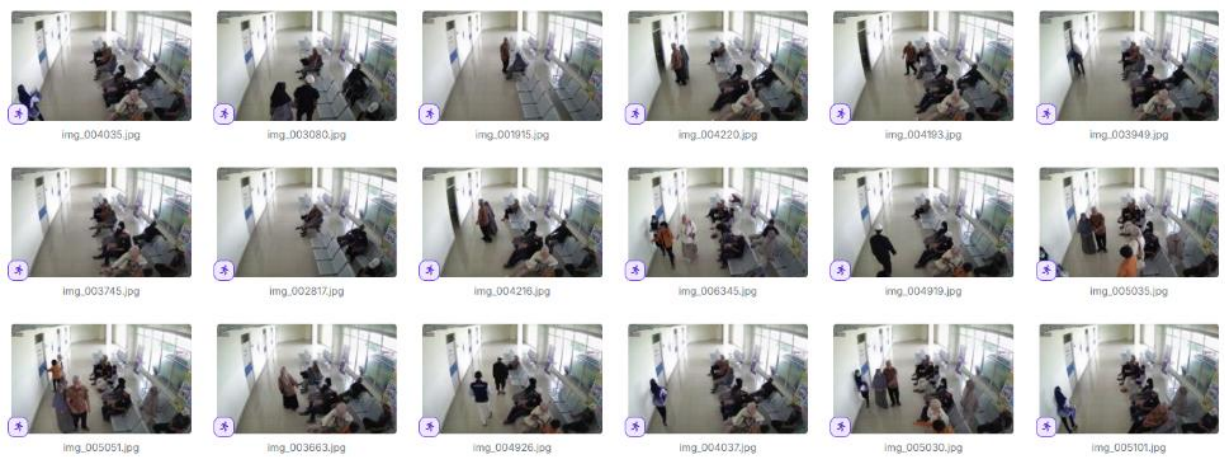
Tahapan menjawab permasalahan yang telah dipaparkan pada bagian sebelumnya, penelitian yang dilakukan dirancang secara sistematis, dimulai dari tahapan pengumpulan data, pre-processing dataset, hingga ekstraksi metrik analitik *spatiotemporal*. Berikut Gambar 1 ilustrasi alur penelitian yang dilakukan. Eksperimen analisis spatiotemporal dilakukan menggunakan komputer dengan prosesor AMD Ryzen 5 5600G with Radeon Graphics dan RAM sebesar 16 GB. Seluruh inferensi model maupun MOT pada perangkat tersebut.



Gambar 1. Alur Penelitian

Pengumpulan data

Tahapan awal yaitu pengumpulan data. Penelitian yang dilakukan memanfaatkan rekaman video cctv yang memuat aktivitas pengunjung di area ruang tunggu. Dataset yang dikumpulkan mencakup kondisi sebenarnya yang terjadi di lapangan serta memiliki kondisi dinamis seperti variasi pencahayaan, tingkat kepadatan pengunjung, hingga oklusi antara pengunjung. Berikut Gambar 2 representasi visual dari kondisi lingkungan ruang tunggu.

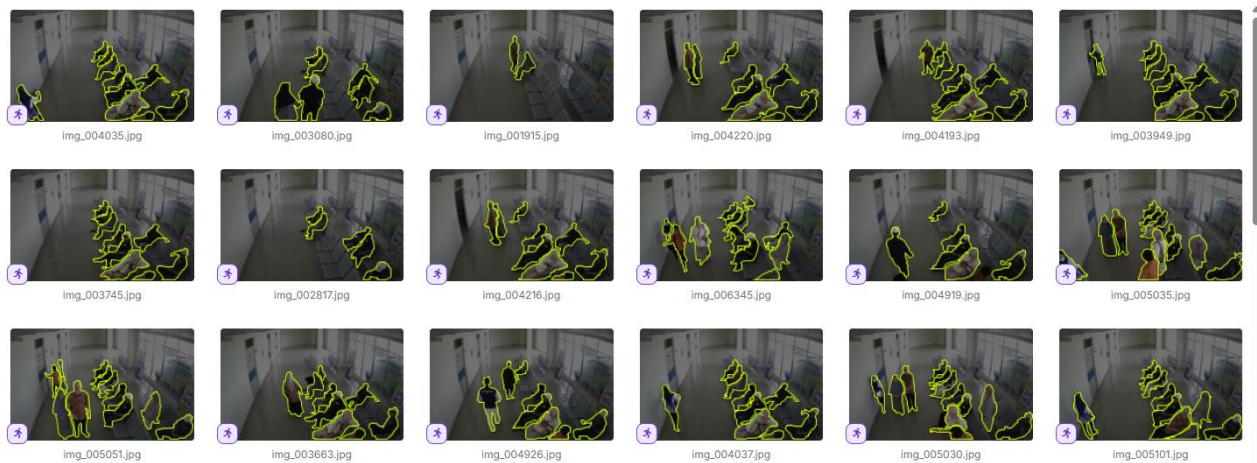


Gambar 2. Kumpulan Dataset

Preprocessing

Tahapan setelah pengumpulan data dilakukan yaitu mengubah video yang telah dikumpulkan menjadi gambar. Perubahan video ke gambar dilakukan dengan memanfaatkan interval setiap 10 Frame menjadi 1 gambar. Meskipun begitu gambar untuk kebutuhan dataset masih ditemukan kemiripan gambar karena interval yang digunakan masih terlalu dekat, sehingga dilakukan proses deteksi kemiripan menggunakan perbandingan ukuran hash dan threshold dari setiap gambar. ukuran hash menggunakan 16 dan *threshold* menggunakan 8, sehingga gambar yang dihasilkan sebanyak 466 gambar. selanjutnya, penyesuaian ukuran gambar 640 x 640 *black edge* [16]. Penggunaan *black edge* pada dataset bertujuan untuk mempertahankan kualitas gambar pada dataset [12], [36].

Dataset yang telah melalui proses sebelumnya, dimasukan ke Roboflow untuk proses anotasi seperti pada Gambar 3 dibawah ini. Hal ini dilakukan untuk memberikan label ke setiap objek yang berada di gambar. penggunaan label atau kelas pada penelitian ini hanya menggunakan binary kelas yaitu: *people*. Anotasi tidak menggunakan *bounding box* namun menggunakan *polygon*. Penggunaan *polygon* karena model yang digunakan nantinya adalah model dengan task segmentasi.



Gambar 3. Dataset yang Telah di Anotasi

Tahapan selanjutnya dalam penelitian yang dilakukan setelah dataset siap yaitu proses mengandakan dataset yang mulanya 466 gambar menjadi 1118 gambar menggunakan teknik augmentasi dataset, namun tidak menurunkan kualitas dari setiap gambar yang ada di dataset. Augmentasi yang dilakukan seperti penyesuaian orientasi gambar ke ukuran 640 x 640 piksel, horizontal flipping, serta variasi nilai hue, saturasi serta kecerahan gambar.

Selanjutnya, hal yang dilakukan adalah melakukan pembagian dataset (*dataset splitting*), pembagian yang dilakukan dengan mengelompokkan menjadi data latih, data validasi, dan data pengujian. Pembagian ini bertujuan untuk melatih model, mengevaluasi performa selama proses pelatihan, serta menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Pelatihan Model

Tahapan pelatihan model ini dilakukan menggunakan Bahasa pemrograman Python versi 3.12.11, PyTorch 2.8.0, CUDA 12.6 yang diakses di platform Google Colab. Penelitian yang dilakukan menggunakan model Instance segmentasi dari YOLOv8n-seg dan YOLOv11n-seg. Semua model yang dilatih menggunakan konfigurasi hypermeter yang sama antara lain: 200 *epochs*, *Learning Rate* (lr) 0.01, momentum 0.937, *weight decay*. Pelatihan model ini juga menggunakan augmentasi dengan konfigurasi antara lain: mosaic 1.0, scale 0.5, parameter perubahan warna HSV terdiri dari hsv h 0.015, hsv s 0.7, dan hsv v 0.4.

Multi-Object Tracking

Model yang telah dilatih sebelumnya akan di integrasikan dengan *Multi-Object Tracking* (MOT). Penggunaan MOT mengadopsi paradigma *tracking-by-detection*. Meskipun begitu pelacakan sepenuhnya bergantung pada kualitas daripada ekstraksi fitur spasial yang dihasil dari model detektor disetiap frame video. Sistem menerima masukan utama berupa koordinat kotak pembatas serta *polygon* segmentasi dari model YOLOv8n-seg dan YOLOv11n-seg, kemudian diproses ke tahapan asosiasi data temporal menggunakan algoritma Deep OC-SORT dan StrongSORT.

Proses pelacakan yang dilakukan memanfaatkan Kalman Filter untuk prediksi posisi lanjutan dari setiap individu yang berada di video. Selain itu, untuk mencapai efisiensi komputasi yang relevan dengan kebutuhan pemantauan *real-time*, proses MOT dioptimalkan menggunakan teknik penyesuaian kecepatan pemrosesan (*frame skipping*), menurunkan analisis pelacakan menjadi 10 hingga 15 *Frame Per Second* (FPS) sehingga beban unit pemrosesan berkurang secara signifikan tanpa mengorbankan akurasi agregasi spasial.

Ekstraksi Informasi Spatiotemporal, Trajectory, Heatmap

Ekstraksi informasi *spatiotemporal* yang dilakukan menggunakan hasil deteksi dari model serta MOT pada setiap frame video. Informasi yang didapatkan berasal dari proses analisis yang meliputi identitas objek (*track ID*), koordinat posisi objek, dan perubahan posisi antar-frame video. Informasi posisi yang digabungkan dengan urutan frame memungkinkan sistem memberi gambaran kondisi spasial pengunjung serta perubahan pergerakan. Hasil ekstraksi yang didapatkan

selanjutnya akan membentuk trajectory, heatmap, proses analisis perubahan jumlah pengunjung (*occupancy*) secara alamiah.

Trajectory yang terbentuk secara alamiah tersebut dapat merepresentasikan lintasan gerak setiap pengunjung. Lintasan tersebut terbentuk berdasarkan perubahan koordinat posisi antar-frame. Setiap objek yang berhasil dilacak diberikan *track ID*, kemudian koordinat centroid dicatat secara berurutan sepanjang proses analisis. Rangkaian koordinat dapat membentuk lintasan pergerakan yang menunjukkan arah dan pola perpindahan setiap pengunjung yang berada di ruang tunggu. Representasi ini digunakan untuk mengidentifikasi pola perpindahan dan aktivitas pengunjung pada area ruang tunggu.

Pembentukan heatmap digunakan untuk merepresentasikan distribusi spasial aktivitas pengunjung berdasarkan akumulasi posisi objek hasil tracking. Setiap titik posisi pengunjung pada frame yang dianalisis diakumulasi ke dalam peta spasial, sehingga akan terlihat area yang sering ditempati atau dilalui. Hasil akumulasi tersebut kemudian di normalisasikan dan divisualisasikan dalam bentuk warna untuk memudahkan identifikasi aktivitas pengunjung. Dengan demikian, heatmap dapat menggambarkan distribusi aktivitas secara spasial dan dapat digunakan untuk mengidentifikasi area dengan tingkat kepadatan pada ruang tunggu.

Metrik Evaluasi

Terdapat dua tahapan untuk melakukan evaluasi, yaitu evaluasi deteksi dan evaluasi estimasi *occupancy*. Kedua tahapan memiliki tujuan antara lain evaluasi deteksi untuk mengukur kemampuan model dalam identifikasi pengunjung, sedangkan evaluasi estimasi *occupancy* untuk mengukur tingkat kesalahan dan kesesuaian hasil prediksi jumlah pengunjung terhadap *ground truth*.

Evaluasi Deteksi

Evaluasi deteksi yang digunakan antara lain *Precision*, *Recall*, *F1-score*, mAP50, mAP50-95. *Precision* bertujuan untuk mengukur ketepatan hasil deteksi, *Recall* bertujuan untuk mengukur kemampuan model dalam menentukan semua objek yang terdapat pada dataset, *F1-score* bertujuan untuk mengukur keseimbangan *precision* dan *recall*. Sementara, mAP50 bertujuan mengukur rata-rata *mean Average Precision* pada tingkat *Intersection over Union* (IoU) sebesar 0.50, sedangkan mAP50-95 sama seperti mAP50 tetapi memiliki rentang IoU 0.5 sampai 0.95. Berikut persamaan *Precision* (1), *Recall* (2), *F1-score* (3), AP (4), mAP (5) [37]. Terkait persamaan AP merupakan persamaan dasar sebelum mendapatkan mAP50 ataupun mAP50-95.

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)} \quad (1)$$

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (TN)} \quad (2)$$

$$F1 - score = 2 * \frac{Precision\ (P) * Recall\ (R)}{Precision\ (P) + Recall\ (R)} \quad (3)$$

$$AP = \int_0^1 p(R) dR \quad (4)$$

$$mAP = \frac{\sum mAP}{m} \quad (5)$$

Evaluasi Estimasi Occupancy

Evaluasi estimasi *occupancy* yang dilakukan menggunakan *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Coefficient of Determination* (R^2). MAE bertujuan untuk mengukur rata-rata selisih absolut antar jumlah pengunjung hasil prediksi dan *ground truth*. RMSE bertujuan untuk memberikan penalty yang lebih besar terhadap kesalahan nilai yang tinggi. [38] bertujuan untuk menunjukkan tingkat kesesuaian variasi hasil estimasi *occupancy* terhadap variasi *ground truth*. Hasil nilai dari MAE dan RMSE rendah menunjukkan estimasi yang lebih baik, sedangkan nilai R^2 yang semakin tinggi menunjukkan hubungan dan kesesuaian hasil estimasi semakin baik [38]. Berikut persamaan MAE (6) [38], RMSE (7), MAPE (8), R^2 (9) [39]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

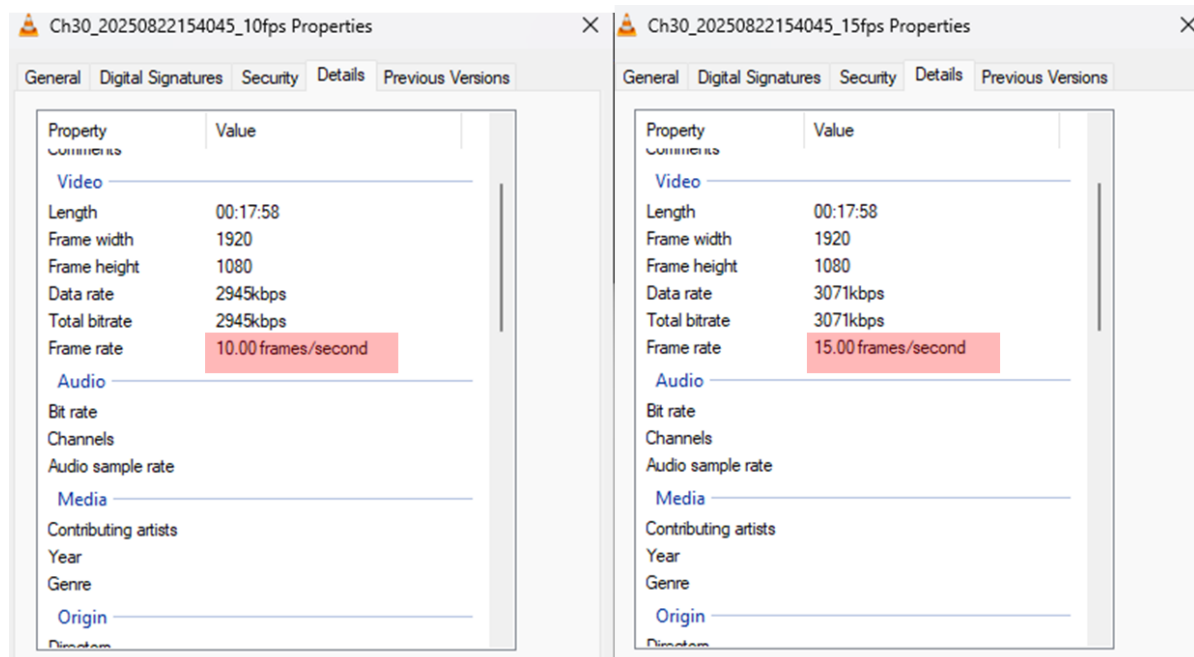
$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

HASIL DAN PEMBAHASAN

Bagian ini menyajikan dari hasil implementasi serta pengujian dalam proses analisis pergerakan pengunjung ruang tunggu berbasis video. Bagian ini mencakup hasil pelatihan, evaluasi model, proses MOT, serta ekstraksi informasi *spatiotemporal* melalui *trajectory*, *heatmap*, dan perubahan *occupancy* pengunjung. Selain itu, hasil estimasi jumlah pengunjung juga dibandingkan dengan ground truth menggunakan MAE, RMSE, MAPE, dan R^2 yang bertujuan mengetahui tingkat kesalahan dan kesesuaian prediksi. Hasil analisis kemudian divisualisasikan melalui dashboard secara real-time untuk merepresentasikan penerapan sistem pada kondisi riil ruang tunggu. Penelitian menggunakan dua skenario 10 dan 15 FPS berdasarkan metadata dari video sumber berikut Gambar 4 merupakan metadata video sumber namun frame rate pemrosesan aktual ditentukan oleh waktu inferensi serta proses tracking pada perangkat yang digunakan, sehingga tidak selalu mencapai frame rate video masukan.

Gambar 4. Metadata video



Hasil Pelatihan Model

Penggunaan *Confusion matrix* (CM) bertujuan untuk melihat distribusi hasil prediksi model terhadap kondisi sebenarnya pada kelas yang digunakan pada evaluasi awal. CM memberikan informasi terkait jumlah *True Positive*, *False Positive*, *False Negative*, dan *True Negative* yang dapat dijadikan dasar perhitungan *Precision*, *Recall*, dan *F1-Score*. Berikut Gambar 5 dan 6 CM dari model yang dilatih.

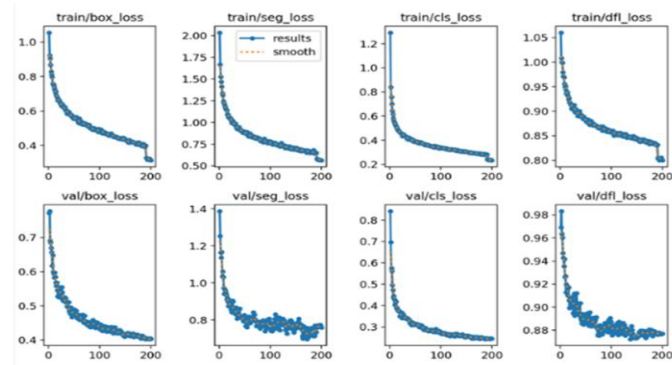


Gambar 5. Confusion Matriks Model YOLOv8n-seg

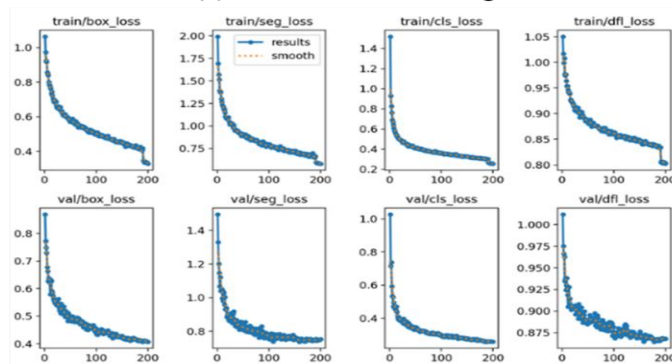


Gambar 6. Confusion Matriks Model YOLOv11n-seg

Selain mengevaluasi hasil dari prediksi berdasarkan kelas, evaluasi juga dilakukan terhadap kualitas segmentasi objek melalui prediksi mask. Evaluasi ini bertujuan untuk melihat kemampuan model yang telah dilatih menghasilkan area segmentasi yang membentuk objek. Berikut Gambar 7 hasil perbandingan performa prediksi mask setiap model.



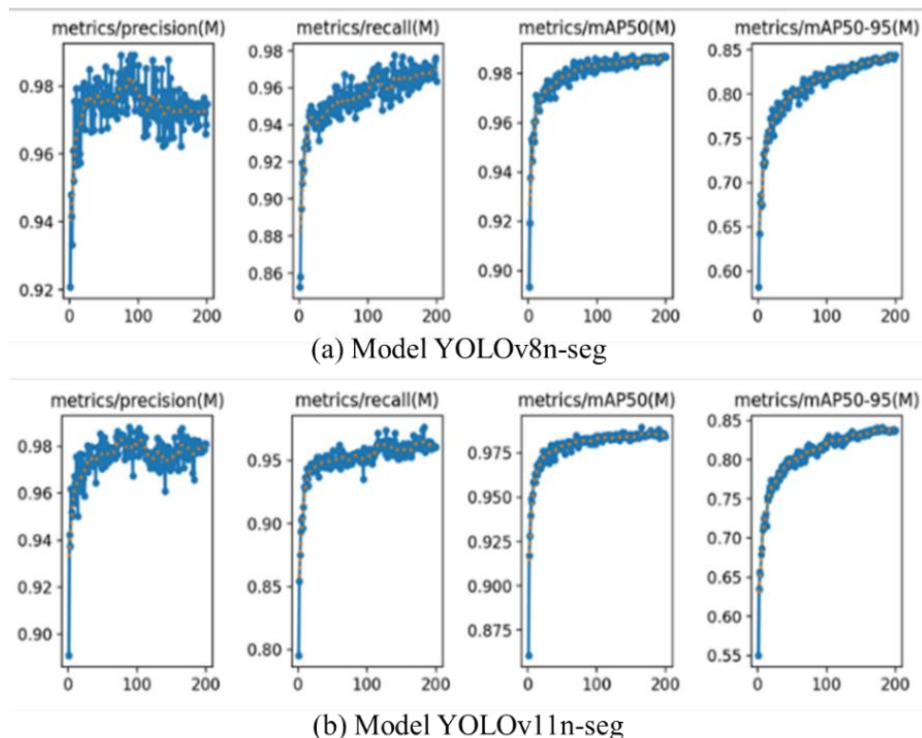
(a) Model YOLOv8n-seg



(b) Model YOLOv11n-seg

Gambar 7. Grafik Perfoma Prediksi Mask (a) YOLOv8n-seg (b) YOLOv11n-seg

Selain evaluasi prediksi *mask*, evaluasi juga dilakukan untuk menganalisis perkembangan *training loss* dan *validasi loss* selama pelatihan model yang berupa grafik. Grafik *training loss* dan *validasi loss* digunakan untuk melihat proses konvergensi model, kestabilan pelatihan, serta perbedaan performa antar pelatihan dan validasi model. Pola perubahan nilai loss dapat menjadi dasar untuk melihat kemampuan model melakukan generalisasi serta mendeteksi kemungkinan model mengalami *overfitting*. Berikut Gambar 8 perkembangan training dan validation loss untuk kedua model.



Gambar 8. Grafik Training dan Validation (a) YOLOv8n-seg (b) YOLOv11n-seg

Berdasarkan hasil evaluasi yang dijelaskan diatas, selanjutnya dilakukan evaluasi menggunakan Precision, Recall, mAP50, mAP50-95, dan F1-Score sebagai indikator kinerja deteksi dan segmentasi objek. Hasil evaluasi kuantitatif kedua model dapat dilihat pada Tabel 2.

Table 2. Hasil Evaluasi Model YOLOv8n-seg dan YOLOv11n-seg

Model	Precision	Recall	mAP50	mAP50-95	F1-score
YOLOv8n-seg	0.972	0.971	0.985	0.844	0.971
YOLOv11n-seg	0.969	0.974	0.986	0.841	0.971

Tabel 2 diatas, menunjukkan kemampuan yang hampir sama dalam mendeteksi dan melakukan segmentasi objek. Model dengan nilai precision yang lebih tinggi menunjukkan kemampuan lebih baik dalam meminimalkan false positive, sedangkan nilai recall yang lebih tinggi menunjukkan model mampu mendeteksi objek yang tersedia pada ground truth. Selanjutnya mAP50 dan mAP50-95 merepresentasikan ketepatan prediksi pada tingkat IoU yang berbeda, sementara F1-Score untuk melihat keseimbangan Precision dan Recall. Berdasarkan keseluruhan metrik, kedua model digunakan untuk tahapan pemrosesan video dan MOT.

Hasil Tracking dan Estimasi Occupancy

Proses tracking dilakukan dengan mempertahankan track ID dari objek yang berada satu frame ke frame berikutnya, sehingga jumlah objek yang berhasil dilacak dapat diperuntukan sebagai estimasi occupancy ruang tunggu. Estimasi tersebut dibandingkan dengan ground truth yang telah ditentukan setiap frame untuk mendapatkan gambaran dari kemampuan sistem dalam presentasi jumlah pengunjung aktual. Pengujian yang dilakukan menggunakan 2 model yaitu: YOLOv8 dan YOLOv11 task model ukuran n segmentasi. Kedua model tersebut diintegrasikan dengan Algoritma Deep OC-SORT dan StrongSORT pada proses analisis video 10 FPS dan 15 FPS. Data hasil estimasi jumlah pengunjung, ground truth, serta kesalahan (*error*) setiap frame digunakan sebagai dasar analisis ke konsistenan tracking dan perubahan

occupancy selama video dianalisis. Berikut Tabel 3 dan Tabel 4 merupakan contoh rincian data hasil estimasi occupancy untuk kedua model.

Table 3. Data Ground Truth dan Hasil Estimasi Jumlah Orang per Frame (YOLOv8n-seg)

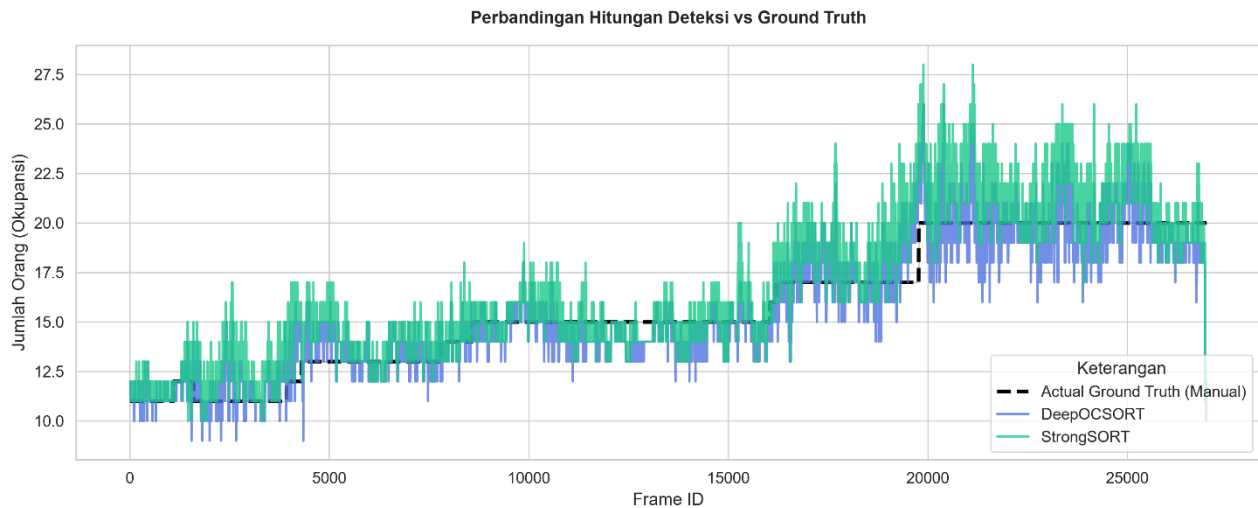
Timestamp	FPS Video	Frame	Tracker	Prediksi	Ground Truth	Absolute Error	Squared Error
6/17/2026 1:02	10	9	Deep OC-SORT	12	11	1	1
6/17/2026 1:02	10	9	StrongSORT	12	11	1	1
6/17/2026 1:02	10	12	Deep OC-SORT	11	11	0	0
6/17/2026 1:02	10	12	StrongSORT	12	11	1	1
....		...	...	...	...	...	...
....		...	...	...	...	...	...
6/17/2026 1:02	15	6	Deep OC-SORT	11	11	0	0
6/17/2026 1:02	15	6	StrongSORT	12	11	1	1
6/17/2026 1:02	15	8	Deep OC-SORT	10	11	1	1
6/17/2026 1:02	15	8	StrongSORT	11	11	0	0

Table 4. Data Ground Truth dan Hasil Estimasi Jumlah Orang per Frame (YOLOv11n-seg)

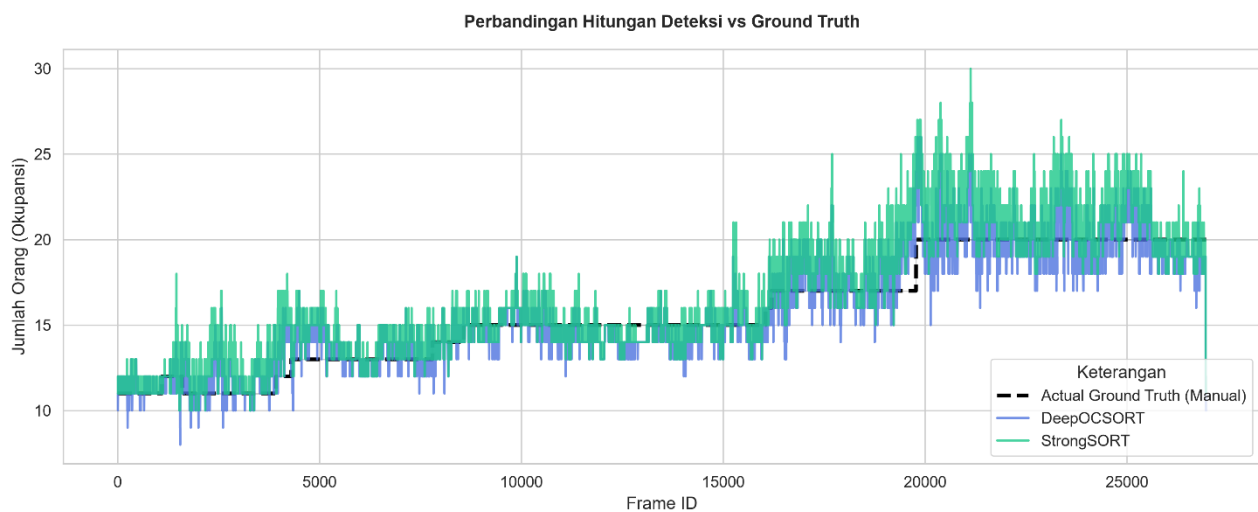
Timestamp	FPS Video	Frame	Tracker	Prediksi	Ground Truth	Absolute Error	Squared Error
6/17/2026 1:02	10	9	Deep OC-SORT	13	11	2	4
6/17/2026 1:02	10	9	StrongSORT	13	11	2	4
6/17/2026 1:02	10	12	Deep OC-SORT	12	11	1	1
6/17/2026 1:02	10	12	StrongSORT	12	11	1	1
....		...	...	...	...	...	...
....		...	...	...	...	...	...
6/17/2026 1:02	15	6	Deep OC-SORT	13	11	2	4
6/17/2026 1:02	15	6	StrongSORT	13	11	2	4
6/17/2026 1:02	15	8	Deep OC-SORT	13	11	2	4
6/17/2026 1:02	15	8	StrongSORT	13	11	2	4

Selain berupa tabel, penelitian ini memvisualisasikan dalam bentuk Grafik pengujian model dan tracker yang dilakukan. Gambar 9 dan Gambar 10, terlihat garis ground truth yang membuktikan adanya perubahan jumlah pengunjung secara bertahap. Kedua algoritma tracker yang digunakan terlihat mengikuti pola perubahan fluktuasi pada beberapa frame. Secara umum, jumlah pengunjung mengalami peningkatan sekitar 11 sampai 12 orang pada awal pengamatan sampai diakhir video yang diujicobakan mencapai 20 orang.

Pola yang tergambarkan pada Gambar 9 dan Gambar 10 menunjukkan bahwa model YOLOv8n-seg yang diintegrasikan dengan Deep OC-SORT dan StrongSORT mampu mengikuti perubahan kondisi okupansi sepanjang rangkaian video. Meskipun begitu ditemukan juga adanya kesalahan dalam perhitungan jumlah orang di beberapa frame. Hal ini menunjukkan bahwa evaluasi tidak hanya melihat nilai occupancy pada satu waktu tertentu, tetapi perubahan jumlah secara temporal masuk kedalam pertimbangan.

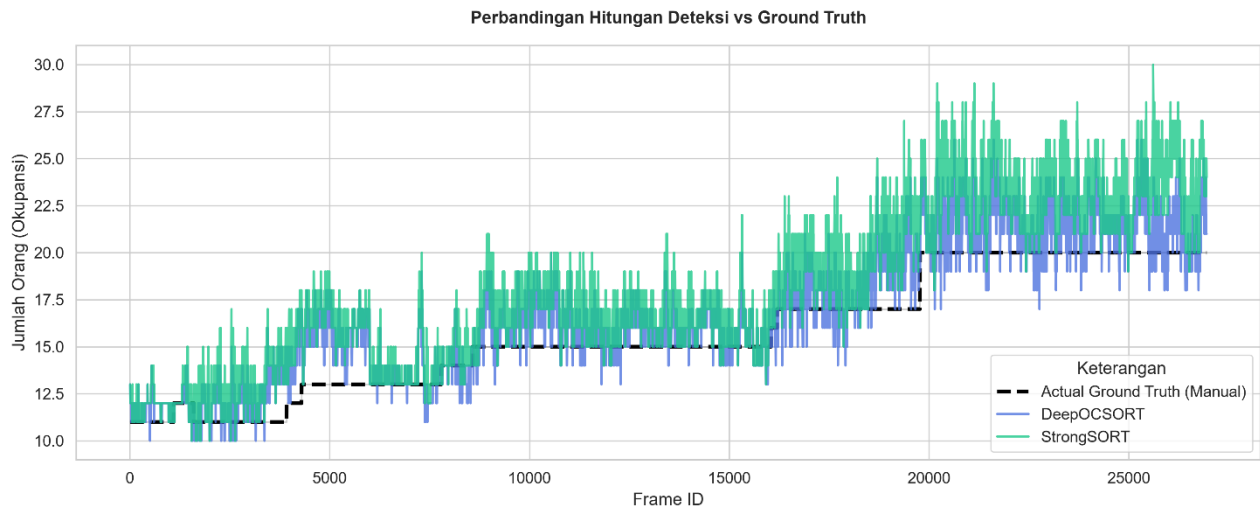


Gambar 9. Perbandingan Prediksi Okupansi terhadap *Ground Truth* (YOLOv8n-seg) 10 FPS

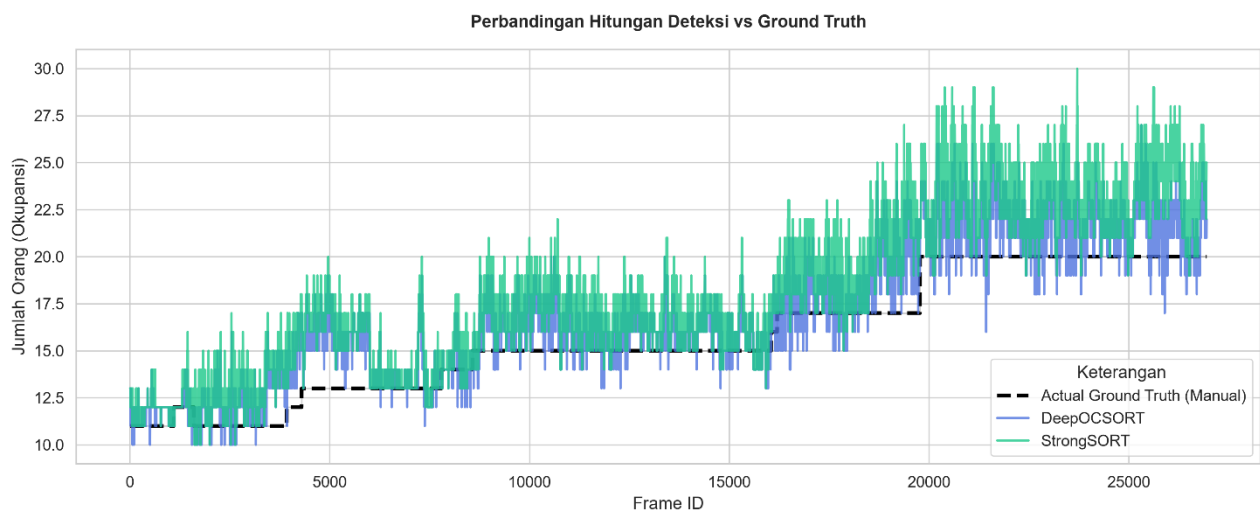


Gambar 10. Perbandingan Prediksi Okupansi terhadap *Ground Truth* (YOLOv8n-seg) 15 FPS

Penggunaan model YOLOv11n-seg yang di integrasikan dengan kedua tracker pada video 10 FPS dan 15 FPS terlihat pada Gambar 11 dan Gambar 12. Kedua grafik merepresentasikan kecenderungan adanya perubahan occupancy yang mengikuti pola ground truth. Meskipun ditemukan adanya fluktuasi prediksi pada kedua tracker. Hal ini menunjukkan bahwa perubahan kepadatan dan kemungkinan oklusi dapat memengaruhi kestabilan estimasi jumlah objek di setiap frame.



Gambar 11. Perbandingan Prediksi Okupansi terhadap *Ground Truth* (YOLOv11n-seg) 10 FPS



Gambar 12. Perbandingan Prediksi Okupansi terhadap *Ground Truth* (YOLOv11n-seg) 15 FPS

Secara keseluruhan, hasil yang telah di jelaskan diatas menunjukkan bahwa output tracking dapat digunakan untuk memberikan gambaran occupancy secara temporal. Informasi tersebut menjadi salah satu komponen penting dalam analisis *spatiotemporal* karena mengintegrasikan hasil pelacakan objek dengan perubahan kondisi kepadatan ruang tunggu dari waktu ke waktu. Selanjutnya, informasi occupancy tersebut dikombinasikan dengan trajectory serta heatmap untuk memperluas analisis dari aspek temporal menuju hubungan spasial dan temporal pergerakan pengunjung.

Evaluasi Error Occupancy

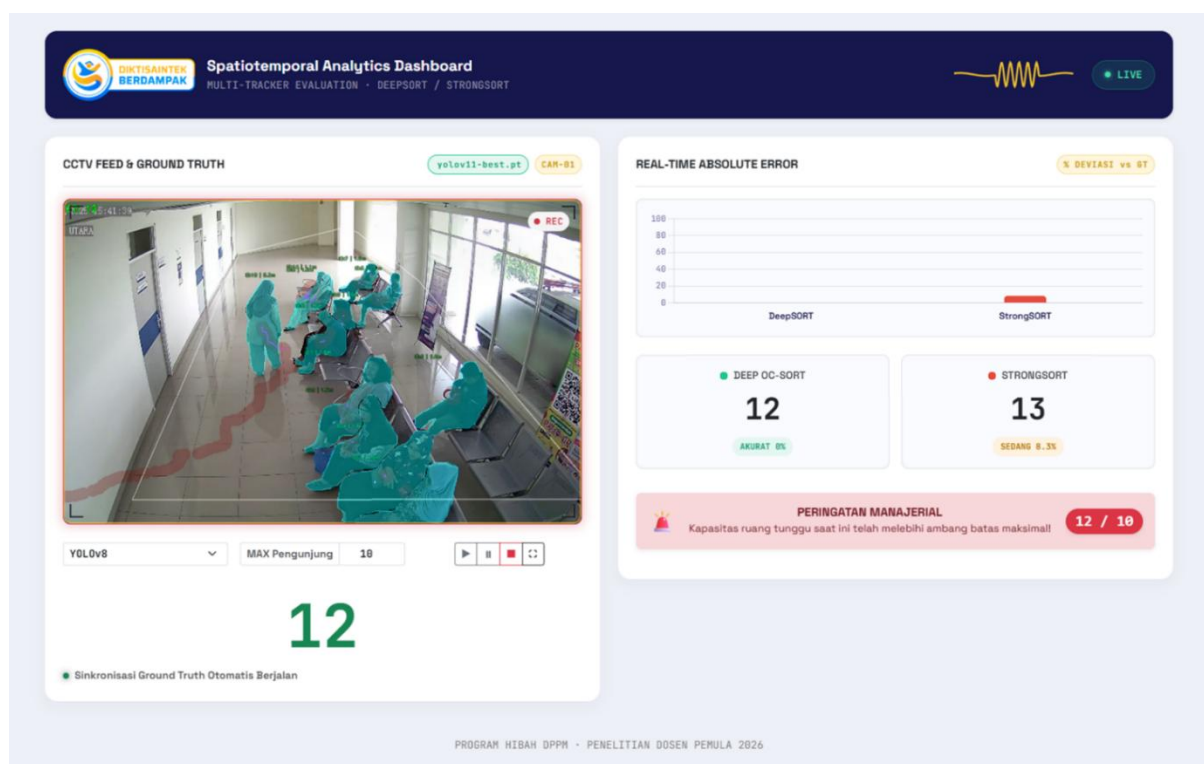
Evaluasi error occupancy dilakukan untuk mengetahui tingkat kesesuaian dari hasil estimasi jumlah pengunjung di kondisi aktual secara keseluruhan. Oleh karena itu, perbandingan antar-FPS dan antar-model perlu dilanjutkan dengan evaluasi kuantitatif menggunakan MAE, RMSE, MAPE, dan R^2 untuk menentukan tingkat kesalahan dan kesesuaian estimasi secara keseluruhan. Evaluasi dilakukan pada kombinasi model dan MOT dengan scenario video 10 FPS dan 15 FPS. Hasil evaluasi error occupancy pengunjung dapat dilihat pada Tabel 5.

Table 5. Evaluasi *Error Occupancy* Pengunjung terhadap *Ground Truth*

Model	MOT	FPS	MAE (orang)	RMSE	MAPE (%)	R ² Score
YOLOv8n-seg	Deep OC-SORT	10	0.94	1.32	5.96	0.8167
		15	0.94	1.31	5.98	0.8218
	StrongSORT	10	1.36	1.89	8.65	0.6258
		15	1.36	1.90	8.70	0.6237
YOLOv11n-seg	Deep OC-SORT	10	1.35	1.79	8.75	0.6638
		15	1.44	1.91	9.35	0.6177
	StrongSORT	10	2.46	3.01	15.48	0.0552
		15	2.49	3.04	15.63	0.0359

Tabel 5 diatas menunjukan kinerja estimasi occupancy terbaik adalah kombinasi antara YOLOv8n-seg dan Deep OC-SORT, dibandingkan kombinasi model dan MOT lainnya. Secara spesifik kombinasi keduanya pada video 10 FPS menghasilkan MAE 0.94, RMSE 1.32, MAPE 5.96% dan R² 0.8167. Selain itu, ditemukan bahwa perubahan frame rate dari 10 FPS ke 15 FPS memberikan perubahan yang relative kecil terhadap tingkat kesalahan estimasi, sementara nilai R² sedikit lebih tinggi pada 15 FPS yang menunjukan kesesuaian variasi *occupancy* yang lebih baik terhadap *ground truth*. Anomali performa terjadi pada model YOLOv11 dengan StrongSORT yang menunjukan rendahnya konsistensi estimasi okupansi. Kondisi tersebut diduga berkaitan dengan sensitivitas proses asosiasi appearance dan motion terhadap occlusion, perubahan pose, kedekatan antarobjek, serta variasi pencahayaan, yang dapat meningkatkan kesalahan asosiasi dan menurunkan keberlanjutan identitas antar-frame. Namun keseluruhan, hasil tersebut menunjukkan bahwa kombinasi model dan tracker memberikan pengaruh terhadap konsistensi estimasi occupancy pada kondisi video yang diuji.

Untuk mendukung proses analisis dan pemantauan hasil deteksi serta pelacakan pengunjung secara langsung, penelitian ini mengembangkan sebuah dashboard analitik *spatiotemporal* yang mengintegrasikan hasil pemrosesan video CCTV dalam satu antarmuka. Dashboard dirancang untuk menampilkan hasil segmentasi dan multi-object tracking, jumlah pengunjung dibandingkan dengan *ground truth*, tingkat kesalahan estimasi, serta informasi spasial dan temporal berupa pergerakan pengunjung dan distribusi aktivitas pada ruang tunggu. Integrasi tersebut memungkinkan hasil analisis dari video tidak hanya digunakan sebagai keluaran algoritma, tetapi juga divisualisasikan secara informatif untuk memudahkan interpretasi kondisi kepadatan dan pola aktivitas pengunjung pada lingkungan ruang tunggu rumah sakit. Tampilan dashboard hasil implementasi penelitian ditunjukkan pada Gambar 13 dan Gambar 14 contoh video yang di analisis secara *real-time*.

Gambar 13. Dashboard *Spatiotemporal* Analitik



Gambar 14. Analisis *Trajectory*, *Heatmap*, *Spatiotemporal*

Berdasarkan Gambar 13 dan Gambar 14 menunjukkan bahwa hasil *tracking* dan estimasi *occupancy* tidak hanya dapat digunakan untuk evaluasi numerik saja, namun dapat diintegrasikan ke dalam media pemantauan operasional secara *real-time*.

Analysis Trajectory

Gambar 13 terlihat adanya penggunaan trajectory untuk memberikan gambaran pola perpindahan pengunjung berdasarkan hasil MOT. jejak pergerakan terlihat dalam bentuk garis berwarna yang mengikuti posisi object dari satu frame ke frame berikutnya. Visualisasi tersebut menunjukkan bahwa sistem tidak hanya mendeteksi keberadaan pengunjung, tetapi juga mempertahankan informasi posisi objek sepanjang rangkaian frame sehingga arah dan pola perpindahan pengunjung dapat di amati. Dengan demikian, trajectory memberikan informasi spasial temporal terkait pergerakan setiap pengunjung di ruang tunggu.

Analisis Heatmap

Selain analisis trajectory pada Gambar 14, visualisasi heatmap yang direpresentasikan melalui area berwarna transparan pada area ruang tunggu. Pada Gambar, aktivitas utama terlihat pada area jalur pergerakan dan sekitar deretan kursi. Dalam konteks pelayanan kesehatan, informasi yang didapatkan dari visualisasi sistem dapat membantu menggambarkan lokasi yang berpotensi mengalami kepadatan.

Analisis Spatiotemporal

Integrasian antara trajectory, heatmap, dan informasi okupansi pada dashboard dapat membentuk analisis *spatiotemporal* pergerakan pengunjung. Aspek spasial ditunjukkan melalui posisi objek, lintasan objek, serta distribusi aktivitas pada heatmap. Sedangkan aspek temporal yang ditunjukkan melalui perubahan jumlah pengunjung berdasarkan proses tracking pada frame video. Gambar 13, menunjukan sistem secara bersamaan menampilkan hasil tracking pengunjung, jumlah pengunjung hasil deteksi, serta ground truth sebagai acuan kondisi aktual.

Integrasi informasi dapat menunjukan bahwa sistem mampu menghubungkan kondisi spasial pengunjung dengan perubahan jumlah pengunjung serta menampilkan tingkat kesalahan estimasi saat analisis oleh model dan tracker. Oleh karena itu, analisis *spatiotemporal* yang dihasilkan tidak hanya menggambarkan dimana pengunjung berada dan bergerak,

tetapi juga menampilkan okupansi ruang tunggu berubah selama video di analisis, sehingga dapat menjadi dasar pemantauan kepadatan dan evaluasi kondisi operasional ruang tunggu secara *real-time*.

KESIMPULAN DAN SARAN

Penelitian yang telah dilakukan ini berhasil mengimplementasikan pendekatan berbasis YOLO dan MOT untuk proses analisis pergerakan dan occupancy pengunjung berdasarkan kondisi aktual di ruang tunggu. Hasil deteksi menggunakan model YOLOv8n-seg dan YOLOv11n-seg selanjutnya diproses menggunakan algoritma pelacakan Deep OC-SORT dan StrongSORT untuk memberikan identitas dan koordinat atau posisi pengunjung pada setiap frame. Informasi tersebut dapat digunakan untuk memberikan estimasi *occupancy*, *trajectory*, *heatmap*, serta analisis perubahan jumlah pengunjung terhadap waktu pada video 10 FPS dan 15 FPS. Evaluasi *occupancy* menggunakan MAE, RMSE, MAPE, dan R^2 menunjukkan adanya perbedaan kinerja antar model dan tracker yang digunakan pada penelitian. Evaluasi pada model Tabel 5 mendapatkan YOLOv8n-seg dengan Deep OC-SORT menghasilkan nilai kesalahan paling rendah dan koefisien determinasi tertinggi, dengan MAE sebesar 0.94 pada 10 maupun 15 FPS, RMSE sebesar 1.32 dan 1.31 MAPE sebesar 5.96% dan 5.98% serta R^2 sebesar 0.8167 dan 0.8218.

Integrasi YOLOv8n-seg dan Deep OC-SORT memberikan kinerja estimasi occupancy paling konsisten pada 2 skenario yang digunakan saat pengujian. Selain itu, integrasi ke dalam dashboard dapat menunjukkan secara real-time terkait kondisi occupancy, hasil tracking, dan indikator kapasitas ruang tunggu. Dengan demikian, penelitian ini menunjukkan bahwa YOLO yang di integrasikan dengan MOT dapat digunakan untuk membentuk informasi spasial dan temporal pengunjung sebagai dasar pemantauan kondisi ruang tunggu.

Meskipun begitu, penelitian selanjutnya disarankan untuk meningkatkan efisiensi proses inferensi serta tracking agar processing latency dapat diminimalkan sehingga informasi temporal, terkhusus dwell time dapat merepresentasikan waktu aktual lebih akurat. Pengujian juga perlu diperluas menggunakan data yang lebih banyak dan bervariasi untuk mengetahui tingkat generalisasi sistem. Selain itu, penelitian selanjutnya dapat mempertimbangkan evaluasi MOT yang lebih komprehensif seperti HOTA, MOTA, serta IDF1.

Selanjutnya, pengembangan *spatiotemporal* berbasis zona diperlukan untuk mengidentifikasi pola kepadatan dan potensi bottleneck yang lebih rinci. Pengembangan lebih lanjut juga dapat diarahkan untuk integrasi data hasil tracking dengan informasi operasional rumah sakit sehingga dashboard tidak hanya berfungsi sebagai media pemantauan, tetapi dapat mendukung keputusan manajerial berbasis *data-driven*.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada Kementerian Pendidikan Tinggi, Sains dan Teknologi Republik Indonesia (Kemdiktisaintek) yang telah mendukung pendanaan penelitian ini melalui Skema Penelitian Dasar Pemula tahun 2026 dengan nomor SK 196/KPA/C3/KPT/2026 dan kontrak 285/C3/DT.05.00/PL-BARU/2026.

DAFTAR PUSTAKA

- [1] A. Salinas-Medina and A. Neme, "Enhancing Hospital Efficiency Through Web-Deployed Object Detection: A YOLOv8-Based Approach for Automating Healthcare Operations," in *2023 Mexican International Conference on Computer Science, ENC 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ENC60556.2023.10508642.
- [2] A. Ardiansyah, A. S. Widagdo, K. N. Qodri, S. F. E. Nugroho, and N. A. R. Putri, "Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT," *JURNAL FASILKOM*, vol. 13, no. 02, pp. 326–333, Sep. 2023, doi: 10.37859/jf.v13i02.5170.
- [3] R. Hasibuan, "Hospital outpatient waiting time and patient satisfaction: an observational comparative study," *Social Medicine*, vol. 18, no. 2, pp. 141–152, May 2025, doi: 10.71164/socialmedicine.v18i2.2025.1611.
- [4] Kementerian Kesehatan, *Keputusan Menteri Kesehatan Republik Indonesia Nomor 129 Tahun 2008 tentang Standar Pelayanan Minimal Rumah Sakit*. Indonesia, 2008.

- [5] Muhammad Takwa, Andi Niartiningsih, Nurul Hidayah Nur, Nurfitriani Nurfitriani, and Mene Paradilla, "Analysis of Waiting Time for Patient Service Using Lean Concept in Outstanding Installations Stella Maris Hospital," *International Journal of Medicine and Health*, vol. 4, no. 1, pp. 14–23, Jan. 2025, doi: 10.55606/ijmh.v4i1.5469.
- [6] F. Y. Sari, R. Rasyid, and F. Firdawati, "Analysis of Waiting Time for Outpatient Services at dr Rasidin Padang Regional Hospital," *Journal La Medihealtico*, vol. 6, no. 2, pp. 477–492, 2025, doi: 10.37899/journallamedihealtico.v6i2.2038.
- [7] W. Afriana, F. Zulyani, and T. Kurniawan, "Application of Lean Six Sigma Method in Hospital Management Process: Performance Optimization and Waste Reduction," *ENDLESS: INTERNATIONAL JOURNAL OF FUTURE STUDIES*, vol. 6, no. 3, pp. 172–184, Nov. 2023, doi: 10.54783/endlessjournal.v6i3.212.
- [8] A. Baligh Jahromi, K. Attarian, A. Asgary, and J. Wu, "Advancing Indoor Epidemiological Surveillance: Integrating Real-Time Object Detection and Spatial Analysis for Precise Contact Rate Analysis and Enhanced Public Health Strategies," *Int. J. Environ. Res. Public Health*, vol. 21, no. 11, p. 1502, 2024, doi: 10.3390/ijerph21111502.
- [9] S. S. Dhanda *et al.*, "Advancement in public health through machine learning: a narrative review of opportunities and ethical considerations," *J. Big Data*, vol. 12, no. 1, p. 154, Jul. 2025, doi: 10.1186/s40537-025-01201-x.
- [10] L. Deng *et al.*, "Hospital crowdedness evaluation and in-hospital resource allocation based on image recognition technology," *Sci. Rep.*, vol. 13, no. 1, p. 299, Jan. 2023, doi: 10.1038/s41598-022-24221-6.
- [11] A. Jain *et al.*, "Improving Patient Flow in a Healthcare Clinic Post COVID-19: A Data Validation and Exploratory Analysis Approach," in *2023 Systems and Information Engineering Design Symposium (SIEDS)*, IEEE, Apr. 2023, pp. 119–124. doi: 10.1109/SIEDS58326.2023.10137848.
- [12] A. Ardiansyah and N. F. Hasan, "Deteksi dan Klasifikasi Penyakit Pada Daun Kopi Menggunakan Yolov7," *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, vol. 12, no. 1, pp. 30–35, 2023, doi: 10.32736/sisfokom.v12i1.1545.
- [13] A. Ardiansyah, A. S. Widagdo, K. N. Qodri, D. Hidayani, and M. Romadhani, "Implementasi Deteksi Tumor Otak Menggunakan YOLOv11 dan Flask," *JURNAL FASILKOM*, vol. 15, no. 2, pp. 415–422, 2025, doi: 10.37859/jf.v15i2.9703.
- [14] A. Ardiansyah, K. N. Qodri, D. Al Banna, and M. Z. Al-Baihaqi, "Pemanfaatan SAM dan YOLOV8 untuk Deteksi dan Segmentation MRI Tumor Otak," *TEKNIMEDIA: Teknologi Informasi dan Multimedia*, vol. 5, no. 1, pp. 82–89, Jun. 2024, doi: 10.46764/teknimedia.v5i1.192.
- [15] M. Madhavan *et al.*, "Object Detection Human Activity Recognition for Improved Patient Mobility and Caregiver Ergonomics," *J. WSCG (Plzen)*, vol. 33, no. 1–2, 2025, doi: 10.24132/JWSCG.2025-2.
- [16] A. Ardiansyah and R. B. Syahputri, "Performance Analysis of YOLOv8, YOLO11, and YOLOE in Detecting Patient Density under Complex Healthcare Conditions," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 14, no. 4, pp. 579–585, 2025, doi: 10.32736/sisfokom.v14i4.2526.
- [17] A. Alam, I. Ullah, and Y.-K. Lee, "Video Big Data Analytics in the Cloud: A Reference Architecture, Survey, Opportunities, and Open Research Issues," *IEEE Access*, vol. 8, pp. 152377–152422, 2020, doi: 10.1109/ACCESS.2020.3017135.
- [18] H. Wang, J. Huang, G. Wang, H. Lu, and W. Wang, "Contactless Patient Care Using Hospital IoT: CCTV-Camera-Based Physiological Monitoring in ICU," *IEEE Internet Things J.*, vol. 11, no. 4, pp. 5781–5797, Feb. 2024, doi: 10.1109/JIOT.2023.3308477.
- [19] G. Hu *et al.*, "Using artificial intelligence and computer vision to detect and monitor unhealthy child-directed food and beverage marketing: a data-driven study," *EclinicalMedicine*, vol. 89, p. 103549, Nov. 2025, doi: 10.1016/j.eclim.2025.103549.
- [20] H. Herfandi, O. S. Sitanggang, M. R. A. Nasution, H. Nguyen, and Y. M. Jang, "Real-Time Patient Indoor Health Monitoring and Location Tracking with Optical Camera Communications on the Internet of Medical Things," *Applied Sciences*, vol. 14, no. 3, p. 1153, Jan. 2024, doi: 10.3390/app14031153.
- [21] C. Li, J. Wang, S. Wang, and Y. Zhang, "A review of IoT applications in healthcare," *Neurocomputing*, vol. 565, p. 127017, Jan. 2024, doi: 10.1016/j.neucom.2023.127017.
- [22] G. V. Orlov and A. N. Kalinichenko, "Computer Vision Model for Analyzing Registry Workload CDC 'Almazov National Medical Research Centre,'" in *2024 V International Conference on Neural Networks and Neurotechnologies (NeuroNT)*, IEEE, Jun. 2024, pp. 58–61. doi: 10.1109/NeuroNT62606.2024.10585604.
- [23] H.-S. Sim and H.-C. Cho, "Enhanced DeepSORT and StrongSORT for Multicattle Tracking With Optimized Detection and Re-Identification," *IEEE Access*, vol. 13, pp. 19353–19364, 2025, doi: 10.1109/ACCESS.2025.3535092.

- [24] H. Lee, D. Kang, H. Park, S. Park, D. Jeong, and J. Paik, "Real-Time Human Group Detection and Clustering in Crowded Environments Using Enhanced Multi-Object Tracking," *IEEE Access*, vol. 12, pp. 184028–184039, 2024, doi: 10.1109/ACCESS.2024.3503661.
- [25] D. A. Maharani, C. Machbub, P. H. Rusmin, and L. Yulianti, "Real-Time Human Tracking Using Multi-Features Visual With CNN-LSTM and Q-Learning," *IEEE Access*, vol. 12, pp. 13233–13247, 2024, doi: 10.1109/ACCESS.2024.3355785.
- [26] E. Prezioso, F. Giampaolo, S. Izzo, M. Savoia, and F. Piccialli, "Integrating Object Detection and Advanced Analytics for Smart City Crowd Management," in *2023 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, IEEE, Oct. 2023, pp. 1–6. doi: 10.1109/ICNSC58704.2023.10318989.
- [27] M. Javaid, A. Haleem, R. P. Singh, R. Suman, S. Khan, and S. Rab, "Adoption of Internet of Things for smart city-enabled smart hospitals," *Intelligent Hospital*, vol. 1, no. 2, p. 100011, Dec. 2025, doi: 10.1016/j.inhs.2025.100011.
- [28] R. Wang, H. Gao, and Y. Liu, "A data fusion-based method for pedestrian detection and flow statistics across different crowd densities," *Journal of Safety Science and Resilience*, vol. 6, no. 1, pp. 105–113, Mar. 2025, doi: 10.1016/j.jnlssr.2024.07.004.
- [29] D. Hidayani, Mustofa Romadhani, and A. Ardiansyah, "Deep Learning Approach for Sleepiness Detection with YOLOv12," *JKTI Jurnal Keilmuan Teknologi Informasi*, vol. 1, no. 1, pp. 5–10, Jun. 2025, doi: 10.61902/jkti.v1i1.1681.
- [30] K. Gkountakos, D. Touska, K. Ioannidis, T. Tsikrika, S. Vrochidis, and I. Kompatsiaris, "Spatio-Temporal Activity Detection and Recognition in Untrimmed Surveillance Videos," in *Proceedings of the 2021 International Conference on Multimedia Retrieval*, New York, NY, USA: ACM, Aug. 2021, pp. 451–455. doi: 10.1145/3460426.3463591.
- [31] S. Tu *et al.*, "Tracking and automatic behavioral analysis of group-housed pigs based on YOLOX+BoT-SORT-slim," *Smart Agricultural Technology*, vol. 9, p. 100566, Dec. 2024, doi: 10.1016/j.atech.2024.100566.
- [32] Y. Li, B. Zhang, Y. Liu, H. Wang, and S. Zhang, "Personnel Monitoring in Shipboard Surveillance Using Improved Multi-Object Detection and Tracking Algorithm," *Sensors*, vol. 24, no. 17, p. 5756, Sep. 2024, doi: 10.3390/s24175756.
- [33] S. John, "Efficient Spatial-Temporal Modeling for Real-Time Video Analysis: A Unified Framework for Action Recognition and Object Tracking," *arXiv preprint arXiv:2507.22421*, 2025.
- [34] Q. Zhang *et al.*, "A Survey of Generative Techniques for Spatial-Temporal Data Mining," *Data Sci. Eng.*, Mar. 2026, doi: 10.1007/s41019-026-00346-w.
- [35] F. I. Hairuddin, S. Azri, and U. Ujang, "Advanced Techniques for Spatio-Temporal Data Management in Graph Databases—A Systematic Review," *Journal of Advances in Information Technology*, vol. 17, no. 1, pp. 14–33, 2026, doi: 10.12720/jait.17.1.14-33.
- [36] G. Fenu and F. M. Mallocci, "DiaMOS plant: A dataset for diagnosis and monitoring plant disease," *Agronomy*, vol. 11, no. 11, p. 2107, 2021.
- [37] A. Ardiansyah, *Buku Monograf Pengenalan Objek Dalam Bidang Kesehatan*. PT. Sonpedia Publishing Indonesia, 2024.
- [38] T. Iida, "Identifying causes of errors between two wave-related data using performance metrics," *Applied Ocean Research*, vol. 148, p. 104024, Jul. 2024, doi: 10.1016/j.apor.2024.104024.
- [39] Y. Zhang, Z. Wang, J. Liu, Y. Li, Z. Huang, and X. Yu, "A metaheuristic-driven categorical boosting framework with interpretability for high-precision prediction of mechanical properties in corroded reinforced concrete beams," *Eng. Appl. Artif. Intell.*, vol. 163, p. 112804, Jan. 2026, doi: 10.1016/j.engappai.2025.112804.

NOMENKLATUR

Persamaan AP.

AP arti dari *Average Precision*

$p(R)$ arti dari *Precision* sebagai fungsi dari *Recall*

dR arti dari perubahan yang sangat kecil pada nilai *Recall*

Persamaan mAP

mAP arti dari *mean Average Precision*

$\sum_m AP$ arti dari *Total Average Precision*

Persamaan MAE, RMSE, MAPE, R2

n arti dari jumlah observasi atau frame yang dievaluasi

y_i arti dari jumlah pengunjung berdasarkan *ground truth* pada observasi ke- i .

$\hat{y}_i$ arti dari jumlah pengunjung hasil estimasi sistem pada observasi ke- i .