

Analisis Perbandingan Metode Yolo Dan *Faster R-CNN* Dalam Deteksi Objek Manusia

DOI: <http://dx.doi.org/10.35889/progresif.v21i2.2890>

Creative Commons License 4.0 (CC BY – NC)



Muhammad Ilham Pratama^{1*}, Nurchim², Eko Purwanto³

^{1,3}Sistem Informasi, Universitas Duta Bangsa Surakarta, Surakarta, Indonesia

²Teknik Informatika, Universitas Duta Bangsa Surakarta, Surakarta, Indonesia

*e-mail *Corresponding Author*: 210101072@mhs.udb.ac.id

Abstract

Human object detection is an important component in surveillance systems, behavior analysis, and crowd management in public spaces such as stadiums, shopping malls, and terminals. However, the detection process often faces obstacles such as inconsistent lighting, complex backgrounds, and high object density. This study aims to compare the performance of two object detection algorithms, namely YOLOv10 and Faster R-CNN, in detecting humans. The dataset used is uniform and covers a wide range of environmental conditions to ensure fair and objective evaluation. This research involves the stages of data collection, pre-processing, model training, testing, and performance evaluation. The test results show that YOLOv10 has a performance advantage with an mAP50 value of 0.75, higher than that of Faster R-CNN which obtained an AP50 of 0.67. Based on these findings, YOLOv10 is recommended for use in applications that require real-time human detection with a high level of accuracy.

Kata kunci: YOLOV10; *Faster R-CNN*; Object Detection

Abstrak

Deteksi objek manusia merupakan komponen penting dalam sistem pengawasan, analisis perilaku, dan pengelolaan keramaian di ruang publik seperti stadion, pusat perbelanjaan, dan terminal. Namun, proses deteksi sering menghadapi kendala seperti pencahayaan yang tidak konsisten, latar belakang kompleks, dan kepadatan objek tinggi. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma deteksi objek, yaitu YOLOv10 dan *Faster R-CNN*, dalam mendeteksi manusia. Dataset yang digunakan bersifat seragam dan mencakup berbagai kondisi lingkungan untuk memastikan evaluasi yang adil dan objektif. Penelitian ini melibatkan tahapan pengumpulan data, pra-pemrosesan, pelatihan model, pengujian, dan evaluasi performa. Hasil pengujian menunjukkan bahwa YOLOv10 memiliki keunggulan performa dengan nilai mAP50 sebesar 0,75, lebih tinggi dibandingkan *Faster R-CNN* yang memperoleh AP50 sebesar 0,67. Berdasarkan temuan tersebut, YOLOv10 direkomendasikan untuk digunakan dalam aplikasi yang membutuhkan deteksi manusia secara real-time dengan tingkat akurasi tinggi.

Kata kunci: YOLOV10; *Faster R-CNN*; Deteksi Objek

1. Pendahuluan

Dalam era digital saat ini, teknologi *computer vision* memainkan peran penting dalam berbagai bidang kehidupan, terutama dalam sistem pengawasan, keamanan, dan analisis perilaku manusia. Salah satu aplikasi utama dari teknologi ini adalah deteksi objek manusia (*human detection*), yang memiliki potensi besar dalam meningkatkan efisiensi pemantauan di area publik seperti pusat perbelanjaan, stadion, terminal, dan fasilitas umum lainnya. Deteksi objek manusia menjadi krusial karena mampu memberikan informasi visual secara otomatis dan *real-time* yang dapat mendukung pengambilan keputusan secara cepat dan akurat[1].

Namun demikian, penerapan sistem deteksi objek manusia masih menghadapi sejumlah tantangan teknis. Salah satu permasalahan utama adalah kebutuhan akan sistem yang mampu bekerja secara *real-time* dengan tingkat akurasi tinggi dalam kondisi lingkungan yang kompleks,

seperti pencahayaan yang rendah, latar belakang yang ramai, atau area padat pengunjung[2] Sistem konvensional seringkali tidak mampu mempertahankan performa yang konsisten dalam berbagai kondisi tersebut. Masalah ini penting untuk ditangani secara kuantitatif, seperti pengukuran akurasi deteksi (*AP - Average Precision*) dan waktu inferensi, yang menjadi indikator kunci dalam keberhasilan sistem deteksi[3]

Untuk mengatasi permasalahan tersebut, pendekatan berbasis *deep learning* telah banyak digunakan karena kemampuannya dalam mengekstraksi fitur visual secara lebih efektif. Dua pendekatan utama yang saat ini digunakan adalah algoritma satu tahap (*one-stage*) seperti YOLO (*You Only Look Once*), dan algoritma dua tahap (*two-stage*) seperti Faster R-CNN. YOLO dikenal efisien dalam pemrosesan *real-time* karena langsung mendeteksi objek tanpa proses proposal wilayah, sehingga sangat sesuai untuk sistem pengawasan langsung[4][5]. Di sisi lain, Faster R-CNN menawarkan akurasi lebih tinggi dengan melalui proses region proposal sebelum klasifikasi, yang menjadikannya lebih stabil dalam lingkungan kompleks atau padat[6]. Berbagai penelitian sebelumnya telah membuktikan keunggulan masing-masing metode, namun belum banyak yang secara spesifik membandingkan kinerja YOLOv10 dan Faster R-CNN dalam konteks deteksi objek manusia pada skenario nyata, seperti pemantauan keramaian publik[7][8].

Penelitian ini bertujuan untuk melakukan analisis komparatif antara algoritma YOLOv10 dan Faster R-CNN dalam mendeteksi objek manusia. Evaluasi dilakukan berdasarkan metrik *AP@0.5* dan *AP@0.5:0.95* guna mengukur akurasi dan performa dari masing-masing metode. Hasil dari penelitian ini diharapkan dapat memberikan panduan yang tepat dalam memilih algoritma deteksi objek yang sesuai dengan kebutuhan spesifik, khususnya dalam aplikasi pemantauan keramaian, distribusi pengunjung, dan pengawasan otomatis di ruang publik.

2. Tinjauan Pustaka

Deteksi objek manusia telah menjadi fokus berbagai penelitian dalam pengembangan sistem pemantauan otomatis, khususnya dalam meningkatkan efisiensi dan akurasi penghitungan di ruang publik maupun sektor ritel. Salah satu penelitian memanfaatkan YOLO untuk mendeteksi dan menghitung jumlah pengunjung secara *real-time* di lingkungan kampus, dengan hasil akurasi mencapai 85% meskipun masih menghadapi tantangan pada objek berdekatan[9]. Sementara itu, penelitian lain mengembangkan sistem peringatan berbasis YOLOv5 untuk toko ritel, yang mampu mengenali kehadiran pengunjung dari rekaman CCTV dan secara otomatis memutar suara sebagai bentuk respons, dengan waktu respons rata-rata antara 0 hingga 0,7 milidetik.[10]

Algoritma YOLO (*You Only Look Once*) telah banyak digunakan dalam berbagai tugas deteksi objek, termasuk dalam bidang medis, seperti klasifikasi lesi kulit, deteksi kelainan retina dan jantung, hingga segmentasi tumor otak [11]. Keunggulan YOLO terletak pada kecepatannya dalam mendeteksi objek secara *real-time* serta akurasinya yang tinggi dibandingkan metode lain, meskipun tetap memiliki keterbatasan seperti kebutuhan dataset yang seimbang dan kebutuhan komputasi yang tinggi. Dalam pengembangan terbarunya, pendekatan seperti YOLO-World memperluas kemampuan YOLO dengan deteksi *open-vocabulary* melalui integrasi visi dan bahasa, sehingga memungkinkan pendeteksian berbagai objek baru tanpa pelatihan ulang[12]. Teknologi ini menunjukkan performa luar biasa baik dalam akurasi maupun kecepatan, serta berpotensi besar untuk diterapkan pada tugas-tugas deteksi lanjutan dan skenario dunia nyata yang lebih kompleks.

Penelitian oleh Jiajia Zheng et al. membandingkan efektivitas algoritma Faster R-CNN dan YOLOv8 dalam mendeteksi gigi *supernumerary* pada radiograf periapikal, dan hasilnya menunjukkan bahwa YOLOv8 memiliki nilai presisi dan akurasi rata-rata lebih tinggi dibandingkan Faster R-CNN serta melampaui kinerja delapan residen kedokteran gigi dalam uji diagnosis[13]. Penelitian oleh Navamani et al. mengembangkan algoritma Faster R-CNN yang ditingkatkan untuk prediksi jenis penyakit kulit pada citra dermoskopik, dan menunjukkan bahwa model ini mampu meningkatkan akurasi, sensitivitas, serta spesifisitas dalam klasifikasi melanoma dan karsinoma sel basal dibandingkan pendekatan sebelumnya[14].

Kebaruan dari penelitian ini terletak pada pendekatan komparatif antara algoritma YOLO dan Faster R-CNN secara khusus dalam konteks deteksi objek manusia, yang dilengkapi dengan pemanfaatan fitur tambahan berupa *heatmaps* untuk visualisasi sebaran objek terdeteksi, serta kalkulasi jumlah individu yang terdeteksi pada tiap citra atau frame. Berbeda dari penelitian sebelumnya yang umumnya hanya menilai akurasi model secara umum, penelitian ini menyajikan analisis mendalam terhadap performa kedua algoritma dalam skenario deteksi manusia secara

real-time, termasuk evaluasi berbasis metrik AP dan implementasi fitur penghitung jumlah orang. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan sistem pemantauan otomatis yang lebih presisi dan informatif, khususnya untuk aplikasi pada area publik dan sistem pengawasan berbasis video.

3. Metodologi

Penelitian ini mengadopsi kerangka kerja CRISP-DM sebagai metode utama dalam proses analisis data. *CrossIndustry Standard Process for Data Mining* Metode ini pertama kali diperkenalkan di tahun 1996[15]. Metode ini menawarkan kerangka kerja sistematis dalam Prosedur eksplorasi data dirancang untuk memberikan arahan strategis dalam menghadapi berbagai tantangan dalam dunia riset maupun kebutuhan industry. *Cross Industry Standard Process for Data Mining* berperan sebagai pendekatan strategis yang bersifat umum dan menjadi standar dalam kegiatan eksplorasi serta analisis data.[16]. Metode CRISP-DM terdiri dari enam tahapan utama, yang dapat diilustrasikan sebagaimana ditampilkan pada Gambar 1 berikut.



Gambar 1. Metode CRISP-DM
(Sumber: Dicoding 2024)

1) Business Understanding

Poin utama dalam bagian ini adalah pemahaman terhadap masalah deteksi objek manusia di ruang publik menggunakan model berbasis *deep learning*. Penelitian bertujuan membandingkan performa YOLOv10 dan Faster R-CNN dalam mendeteksi manusia secara otomatis. Dengan model deteksi yang akurat, sistem pengawasan dan pemantauan keramaian dapat berjalan lebih efisien dan efektif.

2) Data Understanding

Tahap ini difokuskan pada proses pengumpulan dan pemahaman terhadap dataset citra yang digunakan dalam pelatihan model. YOLOv10 dan Faster R-CNN untuk deteksi objek manusia di ruang publik. Dataset yang diambil melalui rekaman CCTV. Gambar-gambar ini kemudian diproses melalui tahap pelabelan (*labeling*) menggunakan platform *Roboflow* untuk memberikan anotasi pada setiap objek manusia

3) Data Preparation

Tahapan melibatkan proses utama, yaitu augmentasi data. Seluruh dataset bersumber dari satu sumber, yaitu rekaman CCTV. Proses *preprocessing* dan *augmentasi* dilakukan menggunakan platform *Roboflow*. Proses prapemrosesan dilakukan guna menyiapkan data sebelum diimplementasikan ke dalam model.

4) Modeling

Tahap modeling berfokus pada penerapan dua pendekatan deteksi objek berbasis *deep learning*, yaitu YOLOv10 dan Faster R-CNN. Kedua model dipilih karena memiliki arsitektur yang berbeda dan performa yang telah terbukti dalam berbagai studi sebelumnya. Pada tahap ini, masing-masing model dilatih menggunakan data yang sama untuk memastikan perbandingan

yang adil. Proses modeling mencakup pelatihan model, validasi, dan pengujian menggunakan data yang telah disiapkan. Pembahasan teknis mengenai parameter pelatihan, arsitektur, dan konfigurasi model akan dijelaskan secara detail pada bab selanjutnya.

5) *Evaluation*

Tahap evaluasi bertujuan untuk menilai performa model deteksi objek yang telah dilatih, yaitu YOLOv10 dan Faster R-CNN, dengan menggunakan data uji. Evaluasi dilakukan berdasarkan metrik-metrik umum yang digunakan dalam sistem deteksi objek, seperti *Precision*, *Recall*, dan *mean Average Precision* (mAP).

6) *Deployment*

Pada tahap akhir, model yang telah selesai dilatih diintegrasikan ke dalam aplikasi berbasis web menggunakan framework Flask. Melalui integrasi ini, pengguna dapat dengan mudah mengunggah video secara langsung melalui antarmuka web, kemudian memperoleh hasil prediksi dari model secara *real-time*.

4. Hasil dan Pembahasan

Penelitian ini menggunakan total 1.211 gambar sebagai dataset utama, yang keseluruhannya bersumber dari rekaman CCTV di ruang publik. Dataset ini telah melalui proses *preprocessing* dan *augmentasi* untuk meningkatkan keragaman data dan mempersiapkannya sebelum digunakan dalam pelatihan model. Setelah proses pembagian, Penelitian ini memanfaatkan data yang terdiri dari 1.062 gambar pada tahap pelatihan, 101 gambar untuk proses validasi, serta 48 gambar yang dialokasikan untuk pengujian model. Model YOLOv10 dilatih selama 15 *epoch* menggunakan data pelatihan dan validasi tersebut dan juga model Faster R-CNN dilatih selama 560 *iteration* menggunakan data pelatihan yang sama. Model ini difokuskan pada deteksi satu kelas objek, yaitu "*Person*".

1) *Business Understanding*

Permasalahan utama yang menjadi landasan penelitian ini adalah kebutuhan akan sistem pendeteksi manusia yang akurat dan efisien di ruang publik, khususnya untuk mendukung sistem pengawasan dan pemantauan keramaian. Dalam praktiknya, kehadiran teknologi deteksi berbasis *deep learning* dapat memberikan solusi otomatis yang lebih andal dibandingkan pemantauan manual, baik dari sisi waktu, biaya, maupun konsistensi kinerja.

Dengan memanfaatkan dua arsitektur model yang berbeda—YOLOv10 dan Faster R-CNN—penelitian ini bertujuan mengevaluasi dan membandingkan kemampuan masing-masing dalam mengenali objek manusia dari rekaman CCTV. Deteksi yang baik akan sangat bermanfaat dalam berbagai skenario dunia nyata.

Pada tahap ini, pembahasan difokuskan pada bagaimana masing-masing model mampu menjawab kebutuhan utama dalam konteks bisnis, yaitu akurasi dalam mendeteksi individu pada lingkungan yang dinamis. Dengan mempertimbangkan faktor-faktor tersebut, hasil dari pengujian model akan dianalisis lebih lanjut untuk mengetahui model mana yang lebih sesuai untuk diimplementasikan dalam sistem pengawasan publik yang membutuhkan deteksi manusia secara otomatis dan akurat.

2) *Data Understanding*

Tahap ini difokuskan pada proses pengumpulan dan pemahaman terhadap dataset citra yang digunakan dalam pelatihan model. YOLOv10 dan Faster R-CNN untuk deteksi objek manusia di ruang publik. Dataset terdiri dari 503 gambar yang diambil melalui rekaman CCTV. Gambar-gambar ini kemudian diproses melalui tahap pelabelan (*labeling*) menggunakan platform *Roboflow* untuk memberikan anotasi pada setiap objek manusia yang terlihat dalam gambar. Tabel 1 menyajikan representasi visual dari dataset yang digunakan, Seluruh dataset hanya terdiri dari satu kelas objek, yaitu *person*.

Table 1. Dataset Gambar

Kelas	Person
Gambar	

3) Data Preparation

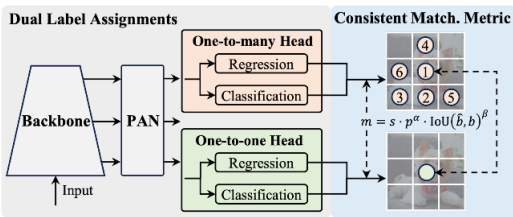
Tahapan melibatkan proses utama, yaitu *augmentasi* data. Seluruh dataset bersumber dari satu sumber, yaitu rekaman CCTV. Proses *preprocessing* dan *augmentasi* dilakukan menggunakan platform *Roboflow*. Proses prapemrosesan dilakukan guna menyiapkan data sebelum diimplementasikan ke dalam model. Sementara itu, proses *augmentasi* dilakukan untuk meningkatkan variasi dataset agar model lebih *robust* terhadap berbagai kondisi. Teknik *augmentasi* yang digunakan meliputi *grayscale*, *rotation*, *saturation*, *blur*, dan *noise*. Setelah *augmentasi*, jumlah dataset meningkat menjadi 1211 gambar. Selanjutnya, Tabel 3 memuat tiga komponen utama, yakni data pelatihan, validasi, dan pengujian. Dataset untuk data training 88% dan 8% untuk data validation, serta data testing sebesar 4%.

Table 2. Pengelompokan Dataset

Kategori	Jumlah	Presentase
Training	1062	88%
Validation	101	8%
Testing	48	4%

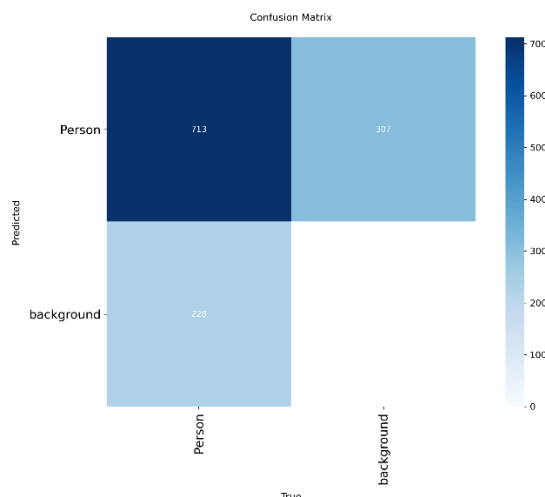
4) Modeling

Model yang digunakan dalam penelitian ini adalah YOLOv10 dan Faster R-CNN. Model YOLOv10 dikembangkan menggunakan bahasa pemrograman Python dengan memanfaatkan library PyTorch melalui framework Ultralytics [17], YOLOv10 merupakan model deteksi objek satu tahap (*one-stage detector*) yang dikenal dengan kecepatan dan efisiensinya dalam melakukan deteksi secara *real-time*.



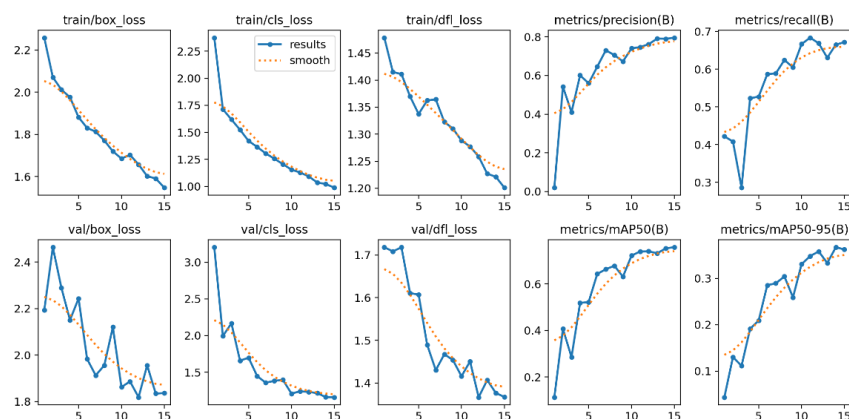
Gambar 2 Architecture Yolo
(Sumber: Yolo 2024)

Hasil pelatihan model divisualisasikan dalam bentuk confusion matrix yang ditampilkan pada Gambar 3. Confusion matrix ini menggambarkan kinerja model dalam membedakan antara objek kelas “*Person*” dan latar belakang (*background*). Dari total 941 sampel yang seharusnya diklasifikasikan sebagai “*Person*” (yang terdiri dari 713 prediksi benar (*true positive*) dan 228 prediksi salah sebagai *background* (*false negative*)), model berhasil mengidentifikasi 713 sampel dengan benar. Namun, sebanyak 228 sampel gagal dikenali, sehingga salah diklasifikasikan sebagai latar belakang. Selain itu, terdapat 307 sampel yang sebenarnya merupakan latar belakang tetapi secara keliru diprediksi sebagai objek “*Person*” (*false positive*).



Gambar 3. Confusion Matrix YOLOv10

Evaluasi performa model dalam proses pelatihan divisualisasikan melalui metrik *mean Average Precision* (mAP) yang ditampilkan pada Gambar 4. Dua metrik utama yang ditinjau adalah $mAP@50$ dan $mAP@50-95$. Grafik $mAP@50$ (*metrics/mAP50(B)*) Hasil pelatihan selama 15 *epoch* menghasilkan nilai $mAP@0.5$ sebesar 0,75 dan $mAP@0.5:0.95$ sebesar 0,36. Nilai ini menunjukkan bahwa model memiliki tingkat akurasi yang baik dalam mendeteksi objek dengan toleransi IoU minimum 50%. mencerminkan kemampuan deteksi objek yang semakin akurat.



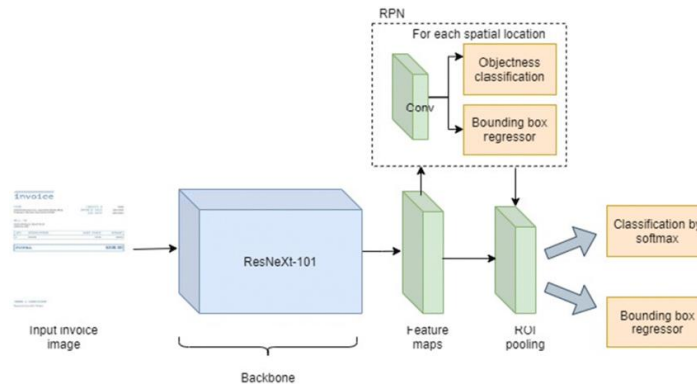
Gambar 4. Hasil Training Model Yolo

Gambar 5. menunjukkan hasil deteksi objek manusia menggunakan model YOLO pada rekaman video pengawasan yang diambil di area publik dengan pencahayaan alami pada siang hari. Dari hasil tersebut, terlihat bahwa model berhasil mengidentifikasi beberapa individu dengan memberikan *bounding box* serta label kelas "Person" disertai *confidence score*.



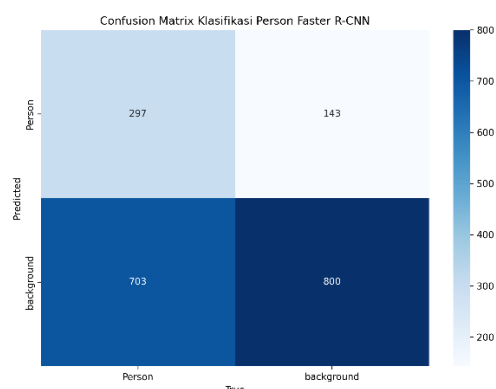
Gambar 5. Hasil Prediksi Model Detection Yolo

Sementara itu, model Faster R-CNN dibangun menggunakan library Detectron2, yang juga berbasis PyTorch dan ditulis dalam bahasa Python. Faster R-CNN merupakan model deteksi objek dua tahap (*two-stage detector*), yang terdiri atas dua komponen utama, yaitu *Region Proposal Network* (RPN) [18], yang berfungsi untuk menghasilkan kandidat wilayah yang berpotensi mengandung objek, dan tahap klasifikasi serta regresi *bounding box* yang dilakukan pada wilayah-wilayah tersebut. Arsitektur Faster R-CNN dalam penelitian ini menggunakan backbone ResNeXt-101 yang didukung oleh Feature Pyramid Network (FPN) guna meningkatkan kemampuan model dalam mengenali objek pada berbagai skala. [19]



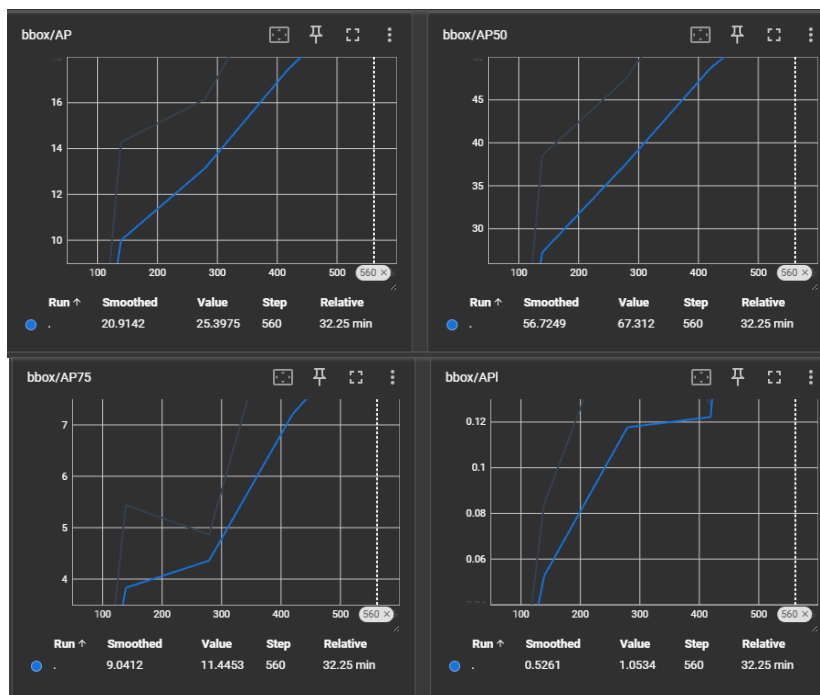
Gambar 6 Architecture Faster R-CNN
(Sumber: researchgate 2022)

Untuk model Faster R-CNN dilatih selama 560 *iteration* menggunakan data *training* dan *validation* yang sama, confusion matrix di tampilkan pada Gambar 7. bahwa dari total 1.000 sampel yang sebenarnya merupakan objek “*Person*”, model hanya mampu mengklasifikasikan Sebanyak 297 data berhasil diklasifikasikan dengan benar (*true positive*), sedangkan 703 data lainnya tidak terdeteksi oleh model dan dianggap sebagai latar belakang (*false negative*). Sementara itu, dari total 943 data yang sebenarnya adalah background, model berhasil mengidentifikasi 800 sampel dengan benar (*true negative*), namun juga melakukan kesalahan klasifikasi terhadap 143 sampel, yang seharusnya merupakan *background* namun diprediksi sebagai objek “*Person*” (*false positive*).



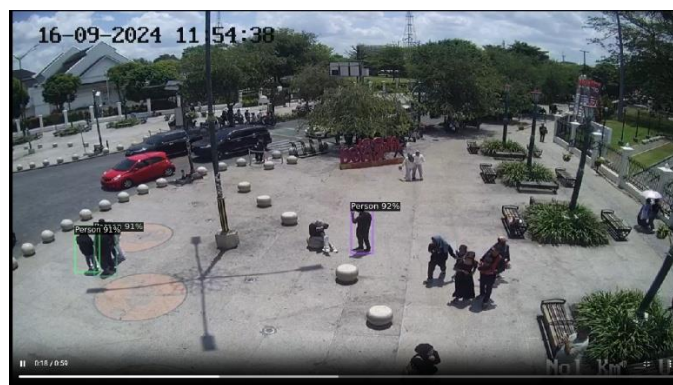
Gambar 7. Confusion Matrix Faster R-CNN

Evaluasi performa model Faster R-CNN selama proses pelatihan divisualisasikan melalui grafik metrik *Average Precision* (AP) pada Gambar 8. Grafik ini meliputi beberapa indikator utama, seperti AP (bbox/AP), AP@50 (bbox/AP50), AP@75 (bbox/AP75), dan area AP (bbox/API), yang masing-masing mencerminkan kualitas prediksi bounding box pada berbagai ambang Intersection over Union (IoU). Nilai ini menunjukkan bahwa model memiliki tingkat akurasi yang baik dalam mendeteksi objek dengan toleransi IoU minimum 50%. Nilai ini merefleksikan tingkat akurasi yang tinggi ketika hanya dibutuhkan tumpang tindih sebagian antara prediksi dan ground truth.



Gambar 8. Hasil Training Model Faster R-CNN

Gambar 9. menunjukkan hasil deteksi objek manusia menggunakan model Faster R-CNN pada data gambar yang sama. Model mampu mengidentifikasi beberapa individu dengan menampilkan *bounding box* dan label kelas "Person". Namun, dibandingkan dengan model YOLO, jumlah objek yang terdeteksi oleh Faster R-CNN lebih sedikit, meskipun setiap deteksi disertai dengan skor kepercayaan yang tinggi.



Gambar 9. Hasil Prediksi Model Faster R-CNN

5) Evaluation

Tabel 3 memperlihatkan hasil evaluasi performa model YOLO dan Faster R-CNN berdasarkan metrik mAP. Kedua model menunjukkan tingkat akurasi yang bervariasi, dengan YOLO mencatat hasil lebih tinggi dibandingkan Faster R-CNN, baik pada mAP@0.5 maupun mAP@0.5:0.95

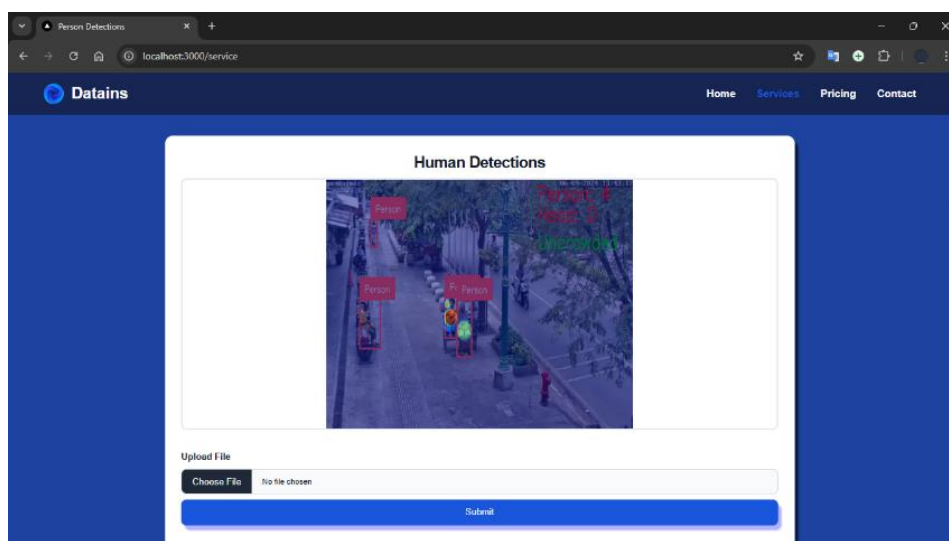
Tabel 3. Hasil Evaluasi Model

Model	Epoch/ Iteration	Precision	Recall	F1-score	mAP@0.5	mAP@0.5:0.95
YOLO	15 Epoch	0.79	0.666	0.72	0.75	0.36
Faster R- CNN	560 Iteration	0.67	0.297	0.41	0.67	0.25

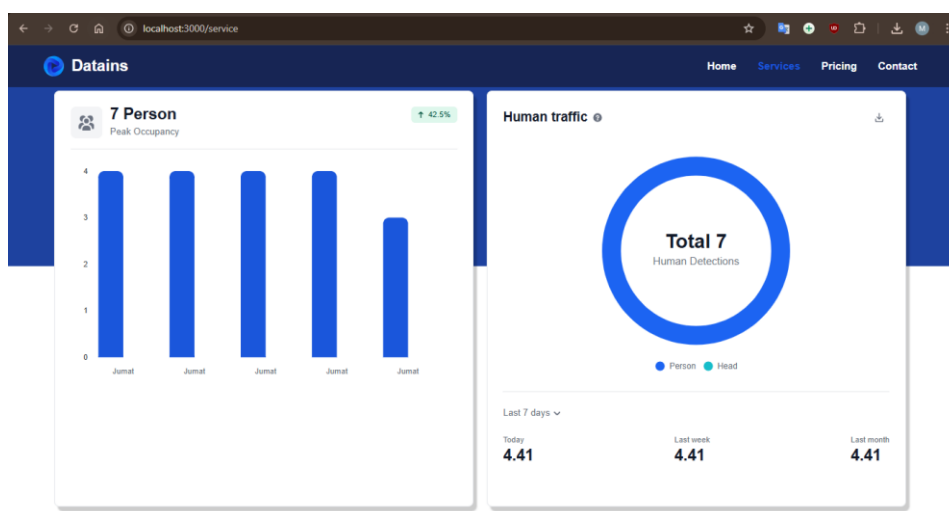
6) Deployment

Setelah proses pelatihan model selesai menggunakan Python, tahap berikutnya adalah melakukan *deployment* ke dalam aplikasi web berbasis Flask. Melalui antarmuka ini, pengguna dapat dengan mudah mengunggah video untuk kemudian diproses secara otomatis oleh sistem. Gambar 10 menunjukkan tampilan halaman unggah yang dirancang sederhana agar pengguna dapat memilih dan mengirim video dari perangkat mereka dengan efisien. Tujuan utama dari *deployment* ini adalah agar model dapat diakses tanpa perlu pemahaman teknis, cukup dengan mengunggah video, maka sistem akan langsung mendeteksi objek (manusia) dan menampilkan hasilnya secara visual.

Selain menampilkan *bounding box* pada objek yang terdeteksi, sistem juga dilengkapi fitur heatmap untuk menunjukkan area dengan kepadatan manusia tertinggi dalam video. Selain itu, sistem secara otomatis menghitung jumlah orang yang terdeteksi dan menyajikannya dalam berbagai bentuk visual seperti *line chart*, *bar chart*, *pie chart*, hingga tabel. Fitur ini memungkinkan pengguna untuk langsung memahami tingkat kepadatan dan distribusi orang dalam video yang diunggah.



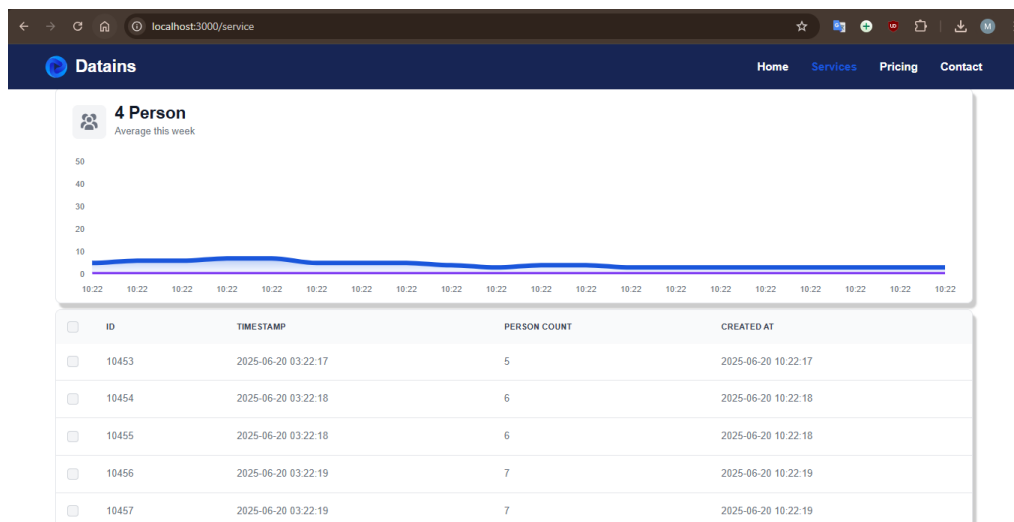
Gambar 10. Tampilan Halaman Web



Gambar 11 Hasil Pemantauan *Dashboard*

Pada halaman dashboard, Gambar 11, pengguna dapat melihat hasil analisis data dalam bentuk bar chart dan pie chart. Grafik ini merepresentasikan jumlah total deteksi objek manusia berdasarkan waktu atau kategori tertentu. Visualisasi ini membantu pengguna memahami pola kepadatan atau aktivitas pengunjung secara umum.

Halaman pemantauan *real-time* Gambar 12. menampilkan line chart yang terus diperbarui secara otomatis (*real-time*) untuk menunjukkan jumlah deteksi manusia dari waktu ke waktu. Selain itu, disediakan pula tabel yang berisi data numerik hasil deteksi, seperti timestamp, jumlah orang per frame, dan ID video. Fitur ini sangat berguna untuk pemantauan langsung pada lokasi, misalnya untuk mengetahui lonjakan jumlah pengunjung dalam periode tertentu.



Gambar 12 Hasil Pemantauan *Real time*

4.3 Pembahasan

Hasil pengujian terhadap dua model deteksi objek manusia, yaitu YOLOv10 dan Faster R-CNN, menunjukkan adanya perbedaan performa yang cukup signifikan. Berdasarkan evaluasi terhadap dataset citra yang digunakan, model YOLOv10 mencatatkan precision sebesar 0.79, recall sebesar 0.666, dan F1-score sebesar 0.7236. Sementara itu, Faster R-CNN hanya mampu mencapai precision sebesar 0.67, recall 0.29, dan F1-score sebesar 0.41. Dari segi mAP@0.5 dan mAP@0.5:0.95, YOLOv10 juga unggul dengan nilai masing-masing 0.75 dan 0.36 dibandingkan Faster R-CNN yang hanya mencatatkan 0.67 dan 0.25.

Perbedaan ini menunjukkan bahwa arsitektur YOLOv10 lebih unggul dalam hal efisiensi dan akurasi deteksi, khususnya dalam konteks *real-time*. Hasil ini sejalan dengan penelitian oleh [20] yang menunjukkan bahwa varian YOLO, termasuk YOLOv10, memiliki waktu inferensi yang jauh lebih cepat dibandingkan Faster R-CNN. Dalam penelitian tersebut, YOLOv10 hanya memerlukan waktu inferensi 19,3 milidetik, sedangkan Faster R-CNN memerlukan 63,8 milidetik. Kecepatan inferensi ini menjadikan YOLO lebih ideal untuk aplikasi yang memerlukan deteksi waktu nyata seperti sistem monitoring dan penghitungan pengunjung.

Secara keseluruhan, hasil penelitian ini memperkuat bahwa YOLO, termasuk versi 10, cenderung lebih unggul dari Faster R-CNN dalam konteks deteksi objek *real-time* dengan akurasi yang relatif tinggi. Faktor arsitektur model yang ringan, pipeline yang end-to-end, serta kemampuan learning spasial yang efisien menjadikan YOLOv10 lebih ideal untuk implementasi praktis di sistem deteksi manusia otomatis.

5. Simpulan

Penelitian ini berhasil merancang dan menguji dua model deteksi objek manusia, yaitu YOLOv10 dan Faster R-CNN, dengan menggunakan dataset manual yang terdiri dari 1.211 citra. Hasil pengujian menunjukkan bahwa YOLOv10 memiliki performa lebih unggul, dengan nilai precision sebesar 0.792. Sementara itu, Faster R-CNN menunjukkan performa yang lebih rendah, dengan precision 0.6757.

Model YOLOv10 juga mencatatkan nilai mAP@0.5 sebesar 0.756 dan mAP@0.5:0.95 sebesar 0.369, dibandingkan Faster R-CNN yang hanya mencapai masing-masing 0.6757 dan 0.2533. Perbedaan performa ini mengindikasikan bahwa YOLOv10 lebih adaptif dan efisien dalam mengenali objek manusia pada lingkungan nyata dengan berbagai variasi citra.

Meskipun hasil yang diperoleh cukup menjanjikan, nilai mAP@0.5:0.95 yang masih di bawah 0.4 menunjukkan bahwa kedua model masih memiliki ruang perbaikan. Faktor-faktor yang

memengaruhi kinerja model antara lain keterbatasan variasi data latih, kondisi pencahayaan yang berbeda, serta kompleksitas latar belakang objek. Untuk meningkatkan performa, disarankan agar dataset diperluas, dilakukan peningkatan jumlah *epoch*, serta penerapan teknik augmentasi yang lebih bervariasi.

Hasil dari penelitian ini diharapkan dapat mendukung pengembangan sistem deteksi manusia otomatis untuk berbagai aplikasi seperti sistem keamanan, pemantauan kerumunan, dan penghitungan pengunjung secara *real-time*.

Daftar Referensi

- [1] B. Putra, G. Pamungkas, B. Nugroho, and F. Anggraeny, "Deteksi dan Menghitung Manusia Menggunakan YOLO-CNN," *J. Inform. dan Sist. Inf.*, vol. 2, no. 1, pp. 67–76, 2021.
- [2] G. Ş. Gündüz dan G. Işık, "A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models," *J. Real-Time Image Process.*, vol. 20, art. no. 5, pp 1-12, Jan. 2023, doi: 10.1007/s11554-023-01276-w.
- [3] D. S. Wirandi, E. D. Permadi, A. Ardiyansyah, D. Prasetyo, M. R. Perani, and R. Rosyani, "Kecerdasan buatan alat pendeteksi maling berbasis Arduino menggunakan sensor ultrasonic melalui SMS," *Scientia Sacra: J. Sains, Teknol. dan Masy.*, vol. 2, no. 2, pp. 841–849, Jun. 2022. [Online]. Available: <http://pijarpemikiran.com/index.php/Scientia>
- [4] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Multimed Tools Appl*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023, doi: 10.1007/s11042-022-13644-y.
- [5] H. Wang, X. Tong, and F. Lu, "Deep Learning Based Target Detection Algorithm for Motion Capture Applications," *J. Phys.: Conf. Ser.*, vol. 1682, no. 1, p. 012032, 2020, doi: 10.1088/1742-6596/1682/1/012032.
- [6] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Jul. 2022, pp. 7464-7475. Available: <http://arxiv.org/abs/2207.02696>
- [7] G. Kucukayan and H. Karacan, "YOLO-IHD: Improved Real-Time Human Detection System for Indoor Drones," *Sensors*, vol. 24, no. 3, art. p. 922, Feb. 2024, doi: 10.3390/s24030922.
- [8] S. V. Akram *et al.*, "Performance analysis of iot and long-range radio-based sensor node and gateway architecture for solid waste management," *Sensors*, vol. 21, no. 8, art. p. 2774, Apr. 2021, doi: 10.3390/s21082774.
- [9] N. Mihuandayani, W. H. S. Pesik, S. Mamuaya, A. N. Putra, and E. Mulyaman, "Implementasi YOLOv8 untuk deteksi dan hitung objek manusia dengan convolutional neural network," *J. Ilm. Informatika (BUFFER)*, vol. 11, no. 1, pp. 27–33, Apr. 2025. [Online]. Available: <https://journal.fkom.uniku.ac.id/index.php/buffer>
- [10] M. A. Mansyur and N. Pratiwi, "Deteksi Manusia Dengan Algoritma YOLO Untuk Pemutaran Audio Otomatis Di Area Tertentu," *JlPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 1, pp. 667–674, Jan. 2025, doi: 10.29100/jipi.v10i1.5967.
- [11] M. G. Ragab *et al.*, "A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023)," *IEEE Access*, vol. 12, pp. 57815–57836, 2024, doi: 10.1109/ACCESS.2024.3386826.
- [12] Y. Xu dan Y. Fu, "Complex Indoor Human Detection with You Only Look Once: An Improved Network Designed for Human Detection in Complex Indoor Scenes," *Appl. Sci.*, vol. 14, no. 22, art. p. 10713, Nov. 2024, doi: 10.3390/app142210713.
- [13] J. Zheng, H. Li, Q. Wen, Y. Fu, J. Wu, and H. Chen, "Artificial intelligent recognition for multiple supernumerary teeth in periapical radiographs based on faster R-CNN and YOLOv8," *J Stomatol Oral Maxillofac Surg*, p. 102293, 2025, doi: 10.1016/j.jormas.2025.102293.
- [14] C. Navamani, S. Sam Daniel, P. Saravanaperumal, R. Thavasi, and S. Venkatesh, "Skin disease type prediction using improved faster R-CNN algorithm," in *Challenges in Information, Communication and Computing Technology*, CRC Press, 2024, pp. 109–114. doi: 10.1201/9781003559092-19.
- [15] L. Purnama and T. Wahyudi, "Analisa Sentimen Tentang Piala Dunia u-20 Indonesia Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Sains dan Teknologi*, vol. 6, no. 2, pp. 217–222, 2024, doi: 10.55338/saintek.v6i2.1397.

-
- [16] Z. Rahman Hakim and S. Sugiyono, "Analisa Sentimen Terhadap Kereta Cepat Jakarta – Bandung Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 939–945, Apr. 2024, doi: 10.55338/saintek.v5i3.1423.
- [17] R. Khanam and M. Hussain, "What is YOLOv5: A deep look into the internal features of the popular object detector," no. 2407, p. 20892, Jul. 2024, [Online]. Available: <http://arxiv.org/abs/2407.20892>
- [18] E. Malagoli dan L. Di Persio, "2D Object Detection: A Survey," *Mathematics*, vol. 13, no. 6, art. p. 893, Mar. 2025, doi: 10.3390/math13060893.
- [19] A. T. Sari, E. Nurlatifah, dan U. Sunan Gunung Djati Bandung, "Penerapan Convolutional Neural Network untuk Mengklasifikasikan Citra Sampah Organik dan Non Organik," *J. Inform.*, vol. 8, no. 2, pp. 120–125, 2024.
- [20] A. Sharma, V. Kumar, dan L. Longchamps, "Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species," *Smart Agric. Technol.*, vol. 9, art. p. 100648, Dec. 2024, doi: 10.1016/j.atech.2024.100648.