

Analisis Perbandingan Antara Model Mask R-CNN dan SSD dalam Mendeteksi Objek Kursi

Elvina Sulisty¹, Fanni Tyasari², Anisa Ismi Azahra³, Dr. Muhammad Munsarif, S.Kom., M.Kom. ⁴

¹Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Semarang, Semarang, Indonesia

²Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Semarang, Semarang, Indonesia

³Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Semarang, Semarang, Indonesia

⁴Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Semarang, Semarang, Indonesia

Info Artikel

Diterima 10 Mei 2025

Perbaikan 13 Juni 2025

Disetujui 28 Juli 2025

Keywords:

Mask R-CNN

Single Short Detection (SSD)

Intersection over Union (IoU)

Model

Score

ABSTRAK

Mendeteksi adalah suatu proses melakukan pemeriksaan terhadap sesuatu objek atau gambar untuk dipelajari atau dipahami dengan menggunakan metode tertentu. Pada penelitian ini dilakukan perbandingan antara model Mask R-CNN dan model *Single Short Detection* (SSD) dalam mendeteksi sebuah objek gambar. Dalam hal ini maka akan diperoleh metode mana yang lebih baik dalam mendeteksi sebuah gambar. Untuk mencapai pada skor yang tinggi maka diperlukan pelatihan dan juga pengujian. Selama pengujian terbukti bahwa antara model Mask R-CNN dan SSD hampir memiliki skor yang sama yaitu mencapai 0.903. yang menjadi perbedaan adalah dibagian Intersection over Union (IoU), dimana dari model Mask R-CNN diperoleh IoU sebesar 0.000 dan model SSD diperoleh IoU sebesar 0.577.

ABSTRACT

Detection is a process of inspecting or examining an object or image to be studied or understood using a specific method. In this research, a comparison is made between the Mask R-CNN model and the Single Short Detection (SSD) model in detecting an image object. In this case, the goal is to determine which method is better in detecting an image. To achieve a high score, training and testing are required. During testing, it was found that both the Mask R-CNN and SSD models almost have the same score, reaching 0.903. The difference lies in the Intersection over Union (IoU) part, where the Mask R-CNN model obtained an IoU of 0.000, and the SSD model obtained an IoU of 0.577.

Ini adalah artikel akses terbuka di bawah lisensi CC BY-SA.



Penulis Korespondensi:

Elvina Sulisty, Fanni Tyasari, Anisa Ismi Azahra

Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Semarang

Alamat: Gedung FT-MIPA Lt. 7, Ruang 707, Jl.Kedungmundu Raya No.18, Semarang 50273, Indonesia

Email: elvinasulisty19@gmail.com , fannityasari@gmail.com , anis yazahra86@gmail.com

1. PENDAHULUAN

Dalam era perkembangan teknologi deteksi objek yang terus berlanjut, sektor-sektor utama seperti keamanan, pemantauan, dan otomasi memusatkan perhatiannya pada penelitian dan pengembangan deteksi objek. Deteksi objek memegang peran sentral dalam pengembangan sistem cerdas dan aplikasi visual yang semakin canggih. Dua model utama yang menjadi sorotan dalam penelitian ini adalah Mask R-CNN dan SSD, masing-masing menawarkan keunggulan uniknya dalam menangani kompleksitas dan variasi objek yang semakin beragam.

Pemilihan dataset Coco Detection Task Object Detection sebagai sumber data penelitian ini menjadi langkah strategis dan penting. Dataset ini memberikan gambaran yang komprehensif tentang beragam objek dalam konteks sehari-hari, menciptakan tantangan signifikan untuk menguji kemampuan kedua model dalam menghadapi variasi bentuk, ukuran, dan tingkat kompleksitas objek. Penelitian ini bertujuan untuk memberikan analisis yang mendalam terkait kinerja Mask R-CNN dan SSD, dengan fokus khusus pada identifikasi objek kursi.

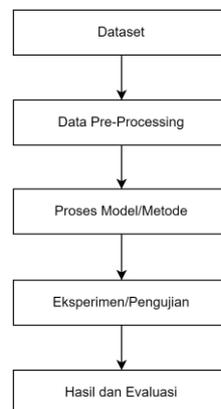
Model Mask R-CNN terkenal akan kemampuannya menghasilkan segmentasi instance dengan tingkat akurasi yang tinggi, memberikan solusi efektif untuk pemetaan objek dengan rincian yang presisi. Di sisi lain, SSD menonjol dengan kecepatan inferensi tinggi, memungkinkan deteksi objek secara real-time tanpa mengorbankan performa. Evaluasi keduanya menggunakan metrik skor dan Intersection over Union (IoU) memberikan wawasan yang lebih mendalam mengenai trade-off antara akurasi dan kecepatan inferensi.

Penelitian ini juga menyoroti relevansi fokus pada identifikasi objek kursi, membawa implikasi praktis yang signifikan untuk berbagai aplikasi, seperti desain ruang, pemantauan publik, dan sektor-sektor lainnya. Dengan memahami karakteristik unik dan keunggulan masing-masing model, diharapkan hasil penelitian ini dapat memberikan panduan berharga bagi para pengembang dan peneliti yang terlibat dalam pengembangan model deteksi objek.

Dalam menghadapi perkembangan teknologi deteksi objek, penting untuk memahami bahwa kesuksesan penerapan model tidak hanya bergantung pada akurasi atau kecepatan semata, tetapi juga pada konteks dan kebutuhan spesifik tugas. Dengan mempertimbangkan secara cermat kelebihan dan kekurangan dari Mask R-CNN dan SSD, para pengambil keputusan dapat membuat pilihan yang tepat sesuai dengan tujuan aplikasi dan persyaratan kinerja yang diinginkan. Dengan demikian, penelitian ini memberikan kontribusi yang substansial pada pemahaman dan perkembangan lebih lanjut dalam bidang deteksi objek, memacu inovasi untuk merancang model-model yang lebih efisien dan adaptif di masa depan.

2. METODE

Pada penelitian ini, tujuan utama adalah membandingkan dua model deteksi objek, yaitu Mask R-CNN dan SSD, dalam konteks mendeteksi objek spesifik, yakni kursi. Metode penelitian ini melibatkan beberapa tahapan yang dirancang dengan cermat untuk memastikan kejelasan dan keberhasilan dalam membandingkan kedua model tersebut.



Gambar 1. Desain Alur Penelitian

Tahap pertama dari penelitian ini melibatkan persiapan dataset. Dataset yang digunakan adalah Coco Detection Task Object Detection, dipilih dengan pertimbangan agar mencakup variasi yang memadai dari objek kursi dalam berbagai konteks sehari-hari. Ketersediaan dataset yang beragam memungkinkan penelitian ini memiliki kekuatan representatif yang tinggi terhadap situasi nyata.

Tahap berikutnya adalah pelatihan model. Dalam tahap ini, model Mask R-CNN dan SSD dilatih menggunakan dataset yang telah dipersiapkan sebelumnya. Pelatihan bertujuan untuk meningkatkan kemampuan model dalam mengenali dan mendeteksi objek kursi. Proses ini mencakup pembelajaran pola dan fitur pada dataset untuk meningkatkan kinerja model.

Setelah pelatihan selesai, tahap evaluasi dilakukan untuk menilai performa kedua model. Objektivitas evaluasi ini ditekankan pada kemampuan model untuk mengidentifikasi dan memetakan objek kursi dengan tingkat akurasi dan kecepatan inferensi yang optimal. Metrik evaluasi yang digunakan melibatkan perhitungan skor dan Intersection over Union (IoU) untuk memberikan gambaran yang komprehensif terkait ketepatan prediksi dan kualitas segmentasi.

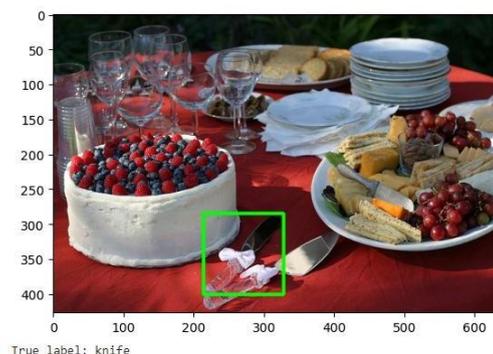
Dengan demikian, metode penelitian ini dirancang untuk mencakup langkah-langkah yang logis dan sistematis, mulai dari persiapan dataset hingga evaluasi performa model. Pendekatan ini diharapkan dapat memberikan pemahaman yang jelas dan menyeluruh mengenai perbandingan antara Mask R-CNN dan SSD dalam konteks deteksi objek kursi.

2.1 Dataset

Data yang diterapkan dalam penelitian ini berasal dari Dataset Coco Detection Task Object Detection, sebuah sumber data yang diakui secara luas dan seringkali digunakan dalam konteks deteksi objek. Keberagaman dan kompleksitas Dataset Coco (Common Objects in Context) membuatnya menjadi pilihan yang sangat berguna, mencakup beragam objek yang sering ditemukan dalam kehidupan sehari-hari. Penelitian ini memiliki tujuan untuk melibatkan analisis kinerja dua model yang terkenal, yaitu Mask R-CNN dan SSD, dalam tugas deteksi objek kursi. Dalam penelitian ini menggunakan 5000 sampel gambar citra yang bisa digunakan untuk uji data tapi dalam hal ini kami mengambil deteksi objek sebuah kursi. 91 label yang diambil menurut kategori yang ditentukan.

2.2 Data Pre Preprocessing

Data Preprocessing adalah langkah awal yang krusial dalam mempersiapkan dataset sebelum masuk ke analisis lebih lanjut. Pada tahap ini, dilakukan transformasi citra untuk mengubah format, ukuran, dan kualitas citra. Dataset kemudian dikelompokkan dan dilabeli, terfokus pada deteksi objek kursi. Pembagian dataset menjadi subset training, testing, dan validation dilakukan untuk evaluasi model. Penghitungan data, termasuk jumlah sampel dan label, dilakukan untuk memahami ukuran dataset. Tahap terakhir melibatkan pembuatan data loader, memastikan penyajian data ke model selama pelatihan dan pengujian. Data Preprocessing menjadi langkah kunci dalam mempersiapkan dataset untuk pemodelan dan evaluasi performa model deteksi objek.

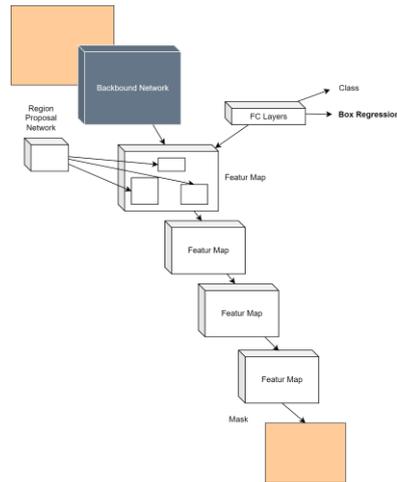


Gambar 2. Labeling gambar

2.3 Proses Model

2.3.1. Arsitektur Model Mask RCNN

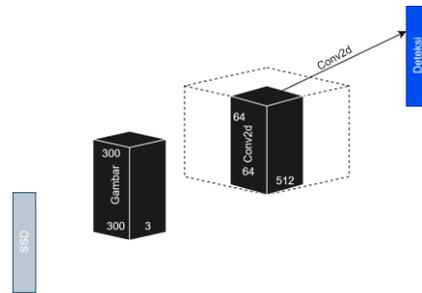
Seperti yang dijelaskan di atas model Mask R-CNN dan SSD digunakan dalam penelitian ini. maka dari itu alangkah baiknya mengerti definisi dari keduanya. Mask R-CNN adalah salah satu algoritma deep learning. Algoritma ini dapat melakukan deteksi objek pada gambar yang sekaligus menghasilkan mask segmentation untuk setiap instance atau biasa disebut dengan instance segmentation. Mask R-CNN disini adalah model yang digunakan untuk mendeteksi gambar atau objek dimana objek yang diangkat adalah sebuah kursi. Mask R-CNN menggunakan RPN untuk menghasilkan proposal region (kotak pembatas) yang kemungkinan mengandung objek. RPN mengidentifikasi area potensial di mana objek dapat berada.



Gambar 3. Arsitektur Model Mask R-CNN

2.3.2. Arsitektur Model SSD

Single Shot Detection (SSD) memiliki pengertian model deteksi yang dapat memproses deteksi objek dengan berbagai metode. Metode yang diterapkan oleh SSD yaitu skala *multi-box* dimana metode ini memungkinkan model untuk mendeteksi objek pada berbagai ukuran dan proporsi. Selain itu, SSD juga menggunakan konvolusi dengan berbagai skala yang memungkinkan untuk mengekstraksi fitur dengan skala yang berbeda. SSD tidak hanya melakukan prediksi terhadap kelas objek, tetapi juga melakukan prediksi terhadap lokasi kotak pembatas. Hal ini memungkinkan SSD untuk secara bersamaan mengidentifikasi objek dan menentukan kotak pembatas yang tepat.



Gambar 4. Arsitektur Model SSD

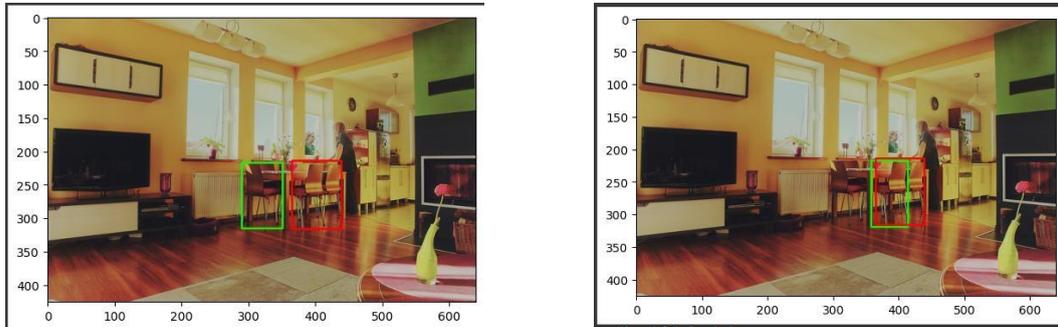
2.4 Pengujian

Dalam upaya melakukan perbandingan antara Model Mask R-CNN dan SSD dalam Mendeteksi Objek Kursi yang melibatkan pengujian metode pengolahan citra (image processing). Pengujian ini dilakukan testing terlebih dahulu dengan memanggil dataset ke 25, 255, dan 4999 sebagai gambar uji, Penelitian ini menggunakan 5000 sampel gambar citra yang bisa digunakan untuk uji data tapi dalam hal ini kami mengambil deteksi objek sebuah kursi dan 91 label yang diambil menurut kategori yang ditentukan. Pada tahap eksperimen, citra diambil dari berbagai sumber, termasuk dataset yang ada. Citra-citra ini kemudian diproses menggunakan teknik image processing yang telah dikembangkan. Pertama-tama, dilakukan segmentasi objek untuk memisahkan elemen elemen deteksi Objek seperti sendok, botol dan pisang. Kemudian, dilakukan ekstraksi fitur untuk mengidentifikasi karakteristik unik dari setiap objek yang terdeteksi. Langkah terakhir adalah klasifikasi, di mana objek-objek ini dikategorikan, misalnya menjadi *chair* atau kursi.

3. HASIL DAN PEMBAHASAN

Dari hasil pelatihan, terlihat bahwa model Mask R-CNN dan SSD menampilkan keunggulan dan kelemahan yang berbeda, mencerminkan trade-off yang harus dipertimbangkan antara akurasi dan kecepatan inferensi, terutama dalam konteks deteksi objek kursi. Kemampuan Mask R-CNN untuk menghasilkan segmentasi piksel yang lebih tepat membuatnya lebih efektif dalam memahami kontur objek kursi, menjadikannya opsi yang menarik untuk tugas-tugas yang membutuhkan akurasi segmentasi tinggi, seperti di bidang medis atau surveilans keamanan. Di sisi lain, SSD menonjol dengan kecepatan inferensi yang tinggi, memungkinkan deteksi objek kursi secara real-time dengan performa yang baik. Meskipun kecepatan ini

krusial dalam beberapa situasi, SSD mungkin perlu mengorbankan sedikit akurasi segmentasi pikselnya. Oleh karena itu, dalam memilih antara kedua model ini, harus diperhitungkan secara cermat kebutuhan spesifik tugas, dengan menemukan keseimbangan yang tepat antara akurasi dan kecepatan inferensi sesuai dengan tujuan penggunaan dan persyaratan performa yang diinginkan.



Gambar 5. Prediksi Model Mask R-CNN (a) Prediksi Model SSD (b)

Tabel 1. Hasil Perbandingan

| Model | Predicted label | Score | IoU |
|------------|-----------------|-------|-------|
| Mask R-CNN | Chair | 0.903 | 0.000 |
| SSD | Chair | 0.903 | 0.577 |

Prediksi model dengan label "chair" tidak hanya mencerminkan kemampuan klasifikasi model, tetapi juga memberikan wawasan tentang sejauh mana model dapat memahami kontur objek kursi. Dengan kemampuan Mask R-CNN untuk menghasilkan segmentasi piksel yang lebih tepat, hasil ini menunjukkan ketangguhan model dalam memahami struktur dan bentuk objek dengan rincian tinggi. Keunggulan ini dapat menjadi kritis, terutama dalam aplikasi yang memerlukan pemetaan objek dengan akurasi tinggi, seperti dalam dunia medis atau keamanan.

Sementara itu, evaluasi performa SSD yang menonjol dalam kecepatan inferensi sejalan dengan skor dan IoU yang diberikan. Meskipun SSD mungkin menghadapi beberapa pengorbanan dalam akurasi segmentasi pikselnya, kecepatan inferensinya memungkinkan deteksi objek kursi secara real-time, menjadi kunci dalam skenario di mana respons cepat sangat diperlukan. Dalam konteks ini, SSD dapat menjadi pilihan yang lebih optimal untuk aplikasi seperti kendaraan otonom atau pemantauan lalu lintas, di mana deteksi objek yang cepat menjadi prioritas utama.

Melalui analisis mendalam terhadap skor keyakinan, IoU, dan karakteristik unik dari setiap model, pengguna dapat mendapatkan pemahaman yang lebih komprehensif. Pemilihan antara Mask R-CNN dan SSD tidak hanya tergantung pada trade-off antara akurasi dan kecepatan inferensi, tetapi juga pada konteks dan kebutuhan spesifik tugas. Dengan pemahaman ini, para pengembang dan peneliti dapat membuat keputusan yang terinformasi dan sesuai dengan persyaratan tugas deteksi objek mereka.

4. KESIMPULAN

Dalam kesimpulan, hasil pelatihan dan evaluasi antara model Mask R-CNN dan SSD memberikan pemahaman yang komprehensif tentang keunggulan dan kelemahan masing-masing dalam tugas deteksi objek kursi. Mask R-CNN, dengan kemampuannya menghasilkan segmentasi piksel yang lebih tepat, menunjukkan performa yang unggul dalam memahami kontur objek, khususnya pada kasus objek kursi. Namun, hasil evaluasi IoU yang lebih rendah menunjukkan adanya potensi peningkatan dalam menentukan batas objek secara lebih akurat.

Di sisi lain, model SSD menonjol dalam hal kecepatan inferensi, memungkinkan deteksi objek kursi secara real-time dengan performa yang baik. Kecepatan ini menjadi kritis dalam skenario yang menuntut respons cepat, seperti pada aplikasi kendaraan otonom atau pemantauan lalu lintas. Meskipun SSD mungkin menghadapi beberapa pengorbanan dalam akurasi segmentasi piksel, evaluasi IoU yang cukup tinggi menunjukkan tingkat kesamaan yang substansial antara area yang diprediksi dan area sebenarnya dari objek kursi.

Dalam memilih antara Mask R-CNN dan SSD, pengguna harus mempertimbangkan dengan matang kebutuhan spesifik tugas mereka. Jika akurasi segmentasi piksel yang tinggi menjadi kunci, terutama dalam

aplikasi medis atau surveilans keamanan, maka Mask R-CNN mungkin menjadi pilihan yang lebih sesuai. Sebaliknya, jika kecepatan inferensi yang tinggi diperlukan, terutama dalam skenario real-time, SSD dapat menjadi solusi yang lebih optimal.

Dengan demikian, hasil evaluasi ini memberikan landasan yang kuat bagi para pengembang dan peneliti untuk membuat keputusan yang terinformasi dalam memilih model deteksi objek sesuai dengan konteks dan persyaratan spesifik tugas mereka. Kesadaran akan keunggulan dan trade-off dari masing-masing model menjadi kunci dalam merancang solusi yang efektif dalam berbagai konteks aplikasi.

REFERENSI

- [1] R. Y. Sabilla and D. Yendri, "Sistem Monitoring Kondisi dan Posisi Pengemudi Berbasis Internet of things," *Chipset*, vol. 2, no. 01, pp. 1–10, 2021, doi: 10.25077/chipset.2.01.1-10.2021.
- [2] V. A. Dihni, "Angka Kecelakaan Lalu Lintas di Indonesia Meningkat di 2021," *databoks.katadata.co.id*, 2022. .
- [3] J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 and Beyond," pp. 1–27, 2023, [Online]. Available: <http://arxiv.org/abs/2304.00501>.
- [4] E. R. Setyaningsih and M. S. Edy, "YOLOv4 dan Mask R-CNN Untuk Deteksi Kerusakan Pada Karung Komoditi," *Teknika*, vol. 11, no. 1, pp. 45–52, 2022, doi: 10.34148/teknika.v11i1.419.
- [5] M. Fathur Rahman Haikal et al., "ANALISIS PERBANDINGAN AKURASI DAN PERFORMA ALGORITMA DETEKSI OBJEK PADA YOLO v3 DENGAN SINGLE SHOT MULTIBOX DETECTOR (SSD)," *Front. Neurosci.*, vol. 14, no. 1, pp. 1–13, 2021.
- [6] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag of freebies sets new state of th art for real-time object detectors," pp. 1–17, 2022, [Online]. Available: <http://arxiv.org/abs/2207.02696>
- [7] L. Cao, X. Zheng, and L. Fang, "The Semantic Segmentation of Standing Tree Images Based on the Yolo V7 Deep Learning Algorithm," 2023.
- [8] S. Zhou et al., "An Accurate Detection Model of Takifugu rubripes Using an Improved YOLO-V7 Network," *J. Mar. Sci. Eng.*, vol. 11, no. 5, 2023, doi: 10.3390/jmse11051051.
- [9] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "Scaled-yolov4: Scaling cross stage partial network," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 13024–13033, 2021, doi: 10.1109/CVPR46437.2021.01283.
- [10] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVgg: Making VGG-style ConvNets Great Again," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, no. 2017, pp. 13728–13737, 2021, doi: 10.1109/CVPR46437.2021.01352.
- [11] R. Wang and J. Li, "Bayes test of precision, recall, and F1 measure for comparison of two natural language processing models," *ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, pp. 4135–4145, 2020, doi: 10.18653/v1/p19-1405.