

Analisis Komparatif YOLO11 dan RT-DETR untuk Deteksi Sampah pada Variasi Pencahayaan

Juni Ismail^{1*}, Pangidoan Adventus Ambarita², Renovand Mikael Situmorang³, Elida Madona Siburian⁴, Ingrid Ester Erlinda Simarmata⁵

¹Teknik Komputer, Politeknik Bisnis Indonesia, Pematang Siantar, Indonesia

^{2,3,4}Sekretari, Politeknik Bisnis Indonesia, Pematang Siantar, Indonesia

⁵Komputerisasi Akuntansi, Politeknik Bisnis Indonesia, Pematang Siantar, Indonesia

Email: ^{1*}juniismaill@gmail.com, ²pangidoanambarita52@gmail.com, ³renovandsitumorang@gmail.com,

⁴madonaely356@gmail.com, ⁵inggridester19@gmail.com

(*Email Corresponding Author: juniismaill@gmail.com)

Received: 25 Juni 2026 | Revision: 30 Juni 2026 | Accepted: 3 Juli 2026

Abstrak

Sistem pemilahan sampah berbasis citra memerlukan model deteksi objek yang mampu mempertahankan akurasi ketika kualitas visual berubah. Penelitian ini menganalisis ketahanan visual YOLO11n, YOLO11s, dan RT-DETR-l pada enam kelas objek sampah, yaitu biodegradable, cardboard, glass, metal, paper, dan plastic. Data eksperimen terdiri atas 7.324 citra latih, 2.098 citra validasi, dan 1.042 citra uji dengan anotasi bounding box berformat YOLO. Evaluasi dilakukan pada test set normal dan empat skenario gangguan, yaitu pencahayaan redup 50%, pencahayaan terang 50%, kontras rendah, dan Gaussian noise. Metrik evaluasi meliputi precision, recall, F1-score, mAP@50, mAP@50:95, dan estimasi FPS. Hasil pengujian normal menunjukkan bahwa RT-DETR-l memperoleh performa tertinggi dengan precision 0,5329, mAP@50 0,4484, dan mAP@50:95 0,3576. Pada evaluasi robustness, RT-DETR-l tetap paling stabil, khususnya pada skenario redup 50% dengan mAP@50 0,4463. Sebaliknya, YOLO11n menghasilkan efisiensi inferensi tertinggi dengan 178,32 FPS pada kondisi normal. Temuan ini menegaskan adanya trade-off antara akurasi deteksi, ketahanan visual, dan kecepatan inferensi untuk sistem deteksi sampah real-time.

Kata Kunci: Deteksi Sampah, YOLO11, RT-DETR, Robustness, Variasi Pencahayaan

Abstract

Image-based waste sorting requires object detection models that can preserve accuracy when visual quality changes. This study analyzes the visual robustness of YOLO11n, YOLO11s, and RT-DETR-l across six waste object classes: biodegradable, cardboard, glass, metal, paper, and plastic. The experimental data consist of 7,324 training images, 2,098 validation images, and 1,042 test images with YOLO-format bounding box annotations. The evaluation was conducted on the normal test set and four disturbance scenarios, namely 50% dimmed lighting, 50% brightened lighting, low contrast, and Gaussian noise. The evaluation metrics include precision, recall, F1-score, mAP@50, mAP@50:95, and estimated FPS. The normal test results show that RT-DETR-l achieved the highest performance with a precision of 0.5329, mAP@50 of 0.4484, and mAP@50:95 of 0.3576. In the robustness evaluation, RT-DETR-l remained the most stable model, particularly under the 50% dimmed scenario with an mAP@50 of 0.4463. In contrast, YOLO11n delivered the highest inference efficiency with 178.32 FPS under normal conditions. These findings confirm a trade-off among detection accuracy, visual robustness, and inference speed for real-time waste detection systems.

Keywords: Waste Detection, YOLO11, RT-DETR, Robustness, Illumination Variation

1. PENDAHULUAN

Peningkatan aktivitas rumah tangga, perdagangan, dan layanan perkotaan berdampak langsung pada bertambahnya volume sampah dengan karakteristik yang semakin beragam. Permasalahan sampah tidak hanya berkaitan dengan kebersihan lingkungan, tetapi juga menyangkut efisiensi pemilahan, pengurangan beban kerja manual, dan kemampuan sistem dalam mengenali jenis material secara konsisten. Pada praktik konvensional, pemilahan sampah masih banyak dilakukan melalui inspeksi manusia. Cara tersebut berpotensi menimbulkan ketidakkonsistenan, membutuhkan waktu relatif lama, dan sulit diterapkan pada skala besar, terutama ketika jenis sampah memiliki bentuk, warna, tekstur, dan kondisi visual yang bervariasi.

Computer vision memberikan alternatif untuk mengotomatisasi proses identifikasi dan pemilahan sampah melalui citra digital. Dalam konteks ini, deteksi objek lebih sesuai dibandingkan klasifikasi citra tunggal karena model tidak hanya menentukan kategori objek, tetapi juga memperkirakan lokasi objek melalui bounding box. Kemampuan tersebut penting karena satu citra dapat berisi beberapa objek sampah dengan ukuran dan kelas berbeda. Namun, implementasi deteksi sampah di lingkungan nyata menghadapi beberapa tantangan, seperti pencahayaan tidak stabil, objek saling tumpang tindih, latar belakang kompleks, kualitas kamera berbeda, serta noise visual. Oleh karena itu, model deteksi sampah perlu diuji tidak hanya pada kondisi data normal, tetapi juga pada skenario gangguan visual yang merepresentasikan kondisi lapangan.

Keluarga You Only Look Once (YOLO) merupakan pendekatan deteksi objek real-time yang banyak digunakan karena memiliki keseimbangan antara akurasi dan kecepatan inferensi [1]. Perkembangan YOLO terus berlanjut melalui peningkatan arsitektur, strategi pelatihan, dan mekanisme optimasi untuk memperkuat performa deteksi [2], [3]. YOLOv9 memperkenalkan Programmable Gradient Information (PGI) dan Generalized Efficient Layer Aggregation Network (GELAN) untuk mengurangi kehilangan informasi selama proses propagasi fitur pada jaringan dalam [4]. YOLOv10 kemudian mengarah pada paradigma end-to-end object detection yang lebih efisien dengan mengurangi ketergantungan terhadap non-maximum suppression [5]. YOLO11 sebagai generasi terbaru dalam ekosistem Ultralytics mendukung berbagai tugas computer vision, termasuk object detection, instance segmentation, classification, pose estimation, dan oriented object detection [6].

Selain pendekatan YOLO, arsitektur berbasis transformer menjadi arah penting dalam object detection modern. Detection Transformer (DETR) merumuskan deteksi objek sebagai permasalahan set prediction secara end-to-end [7]. Meskipun DETR awal memiliki tantangan konvergensi dan beban komputasi, RT-DETR dikembangkan sebagai Real-Time Detection Transformer dengan efficient hybrid encoder dan uncertainty-minimal query selection untuk memperbaiki keseimbangan antara akurasi dan kecepatan [8]. Penelitian RT-DETR melaporkan bahwa varian R50 dan R101 mampu mencapai 53,1% dan 54,3% AP pada COCO serta 108 dan 74 FPS pada GPU T4 [8]. Fakta ini menjadikan RT-DETR relevan untuk dibandingkan dengan model YOLO pada skenario deteksi sampah real-time.

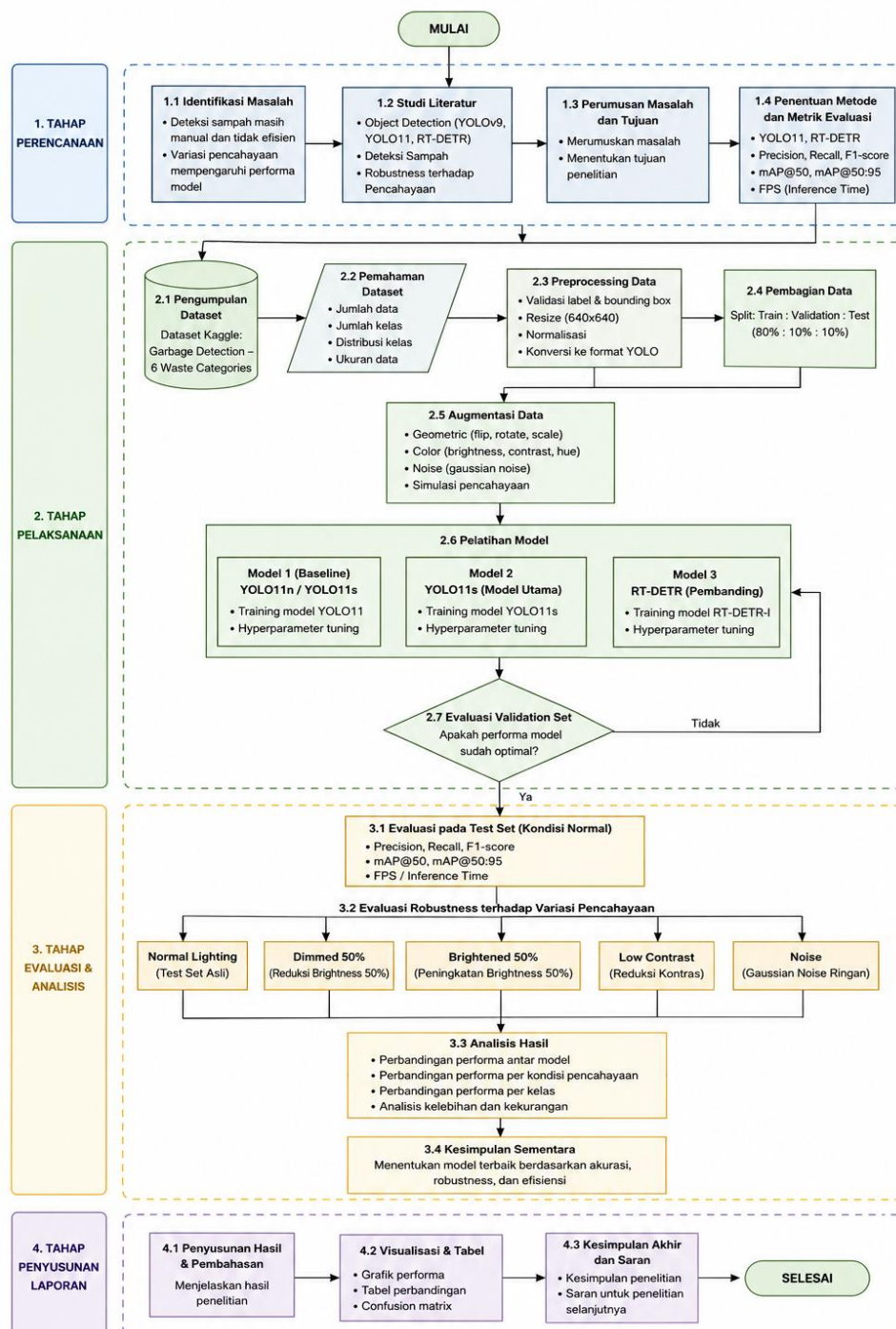
Penelitian terkait deteksi objek dan pengenalan sampah telah menunjukkan bahwa deep learning dapat membantu proses identifikasi material secara otomatis [15]. Namun, sebagian penelitian masih berfokus pada evaluasi performa normal dan belum secara khusus mengkaji stabilitas model ketika kualitas visual berubah. Padahal, sistem deteksi sampah yang ditempatkan pada ruang publik, lingkungan semi-terbuka, atau perangkat edge dapat menghadapi pencahayaan redup, overexposure, kontras rendah, dan noise sensor. Evaluasi robustness menjadi penting agar model yang dipilih tidak hanya memiliki nilai mAP tinggi, tetapi juga mampu mempertahankan performa pada kondisi visual yang tidak ideal.

Berdasarkan permasalahan tersebut, penelitian ini mengevaluasi YOLO11n, YOLO11s, dan RT-DETR-1 pada deteksi sampah real-time. Kontribusi penelitian ini meliputi: pertama, membandingkan model ringan, model ukuran kecil, dan model berbasis transformer pada enam kelas sampah; kedua, menganalisis distribusi objek dan potensi class imbalance; ketiga, mengevaluasi performa normal menggunakan precision, recall, F1-score, mAP@50, dan mAP@50:95; keempat, menguji robustness model terhadap variasi pencahayaan dan noise; serta kelima, membahas trade-off antara akurasi, robustness, dan FPS. Dengan demikian, penelitian ini diharapkan dapat memberikan dasar pemilihan model yang lebih tepat untuk sistem deteksi sampah otomatis.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan metode eksperimen kuantitatif dengan membandingkan performa tiga model deteksi objek pada data sampah beranotasi bounding box. Rancangan penelitian dibuat agar proses pelatihan dan evaluasi dapat direplikasi, mulai dari pengumpulan data, validasi format anotasi, analisis distribusi kelas, pelatihan model, evaluasi normal, hingga pengujian robustness. Alur penelitian disajikan pada Gambar 1.



Gambar 1. Diagram alir tahapan penelitian

Gambar 1 memperlihatkan empat tahap utama penelitian, yaitu perencanaan, pelaksanaan, evaluasi dan analisis, serta penyusunan laporan. Tahap perencanaan mencakup identifikasi masalah, studi literatur, perumusan tujuan, dan pemilihan metode. Tahap pelaksanaan mencakup pengumpulan data, pemahaman dataset, preprocessing, pembagian data, augmentasi, dan pelatihan model. Tahap evaluasi mengukur performa pada kondisi normal serta variasi pencahayaan. Tahap akhir menyusun interpretasi hasil, visualisasi, kesimpulan, dan saran penelitian lanjutan.

2.2 Dataset dan Preprocessing

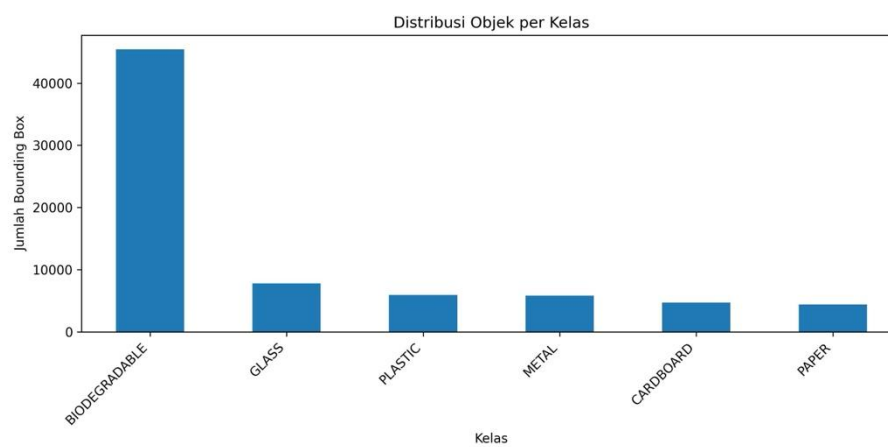
Dataset yang digunakan merupakan kumpulan citra sampah berformat YOLO dengan enam kelas objek, yaitu BIODEGRADABLE, CARDBOARD, GLASS, METAL, PAPER, dan PLASTIC. Setiap label berisi lima komponen utama, yaitu class id, koordinat pusat x, koordinat pusat y, lebar bounding box, dan tinggi bounding box yang dinormalisasi pada rentang 0 sampai 1. Struktur data terdiri atas folder train, validation, dan test, sehingga dapat langsung digunakan pada pipeline Ultralytics.

Preprocessing dilakukan melalui validasi pasangan citra dan label, pemeriksaan format anotasi, penyesuaian konfigurasi kelas, serta penetapan ukuran input model. Augmentasi visual digunakan untuk memperluas variasi data pelatihan, sedangkan dataset robustness dibuat dari test set dengan mempertahankan label bounding box yang sama. Jumlah data pada setiap split ditampilkan pada Tabel 1.

Tabel 1. Ringkasan pembagian dataset

Split	Jumlah Citra	Jumlah Label	Fungsi
Train	7.324	7.324	Pelatihan model
Validation	2.098	2.098	Pemantauan performa selama training
Test	1.042	1.042	Evaluasi akhir dan robustness

Distribusi objek dianalisis berdasarkan jumlah bounding box, bukan hanya jumlah citra. Pendekatan ini lebih sesuai untuk object detection karena satu citra dapat memiliki lebih dari satu objek. Hasil analisis menunjukkan bahwa kelas BIODEGRADABLE jauh lebih dominan dibandingkan kelas lain, sehingga terdapat indikasi class imbalance. Distribusi objek per kelas disajikan pada Gambar 2.



Gambar 2. Distribusi objek per kelas berdasarkan jumlah bounding box

2.3 Model dan Konfigurasi Training

Tiga model digunakan dalam penelitian ini untuk merepresentasikan perbedaan kompleksitas arsitektur. YOLO11n digunakan sebagai baseline ringan yang mengutamakan kecepatan inferensi. YOLO11s digunakan sebagai model utama dengan kapasitas lebih tinggi dibandingkan YOLO11n. RT-DETR-l digunakan sebagai pembanding berbasis transformer yang memiliki kemampuan pemodelan konteks global. Pelatihan dilakukan menggunakan Python 3.12.13, PyTorch 2.10.0+cu128, Ultralytics 8.4.84, dan GPU Tesla T4. Konfigurasi eksperimen diringkas pada Tabel 2.

Tabel 2. Konfigurasi model dan eksperimen

Model	Peran	Epoch	Image Size	Batch
YOLO11n	Baseline ringan	30	640 x 640	16
YOLO11s	Model utama	50	640 x 640	16
RT-DETR-l	Pembanding transformer	20	640 x 640	8

2.4 Metrik Evaluasi

Evaluasi dilakukan menggunakan metrik yang umum digunakan pada object detection, yaitu Intersection over Union (IoU), precision, recall, F1-score, average precision (AP), mean average precision (mAP), dan frame per second (FPS). IoU menghitung tingkat tumpang tindih antara bounding box prediksi dan ground truth. Precision menunjukkan proporsi prediksi positif yang benar, sedangkan recall menunjukkan proporsi objek aktual yang berhasil ditemukan. F1-score digunakan untuk menyeimbangkan precision dan recall. AP dan mAP digunakan untuk mengukur kualitas deteksi pada berbagai ambang keyakinan, sedangkan FPS digunakan untuk menilai efisiensi inferensi. Rumus evaluasi dituliskan pada Persamaan (1) sampai (8).

$$IoU = \frac{Area(B_p \cap B_{gt})}{Area(B_p \cup B_{gt})} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \quad (4)$$

$$AP_k = \int_0^1 P_k(R) dR \quad (5)$$

$$mAP = \frac{1}{C} \sum_{k=1}^C AP_k \quad (6)$$

$$FPS = \frac{1000}{t_{pre} + t_{inf} + t_{post}} \quad (7)$$

$$D(\%) = \frac{M_n - M_c}{M_n} \times 100\% \quad (8)$$

Pada persamaan tersebut, TP adalah true positive, FP adalah false positive, FN adalah false negative, C adalah jumlah kelas, P adalah precision, R adalah recall, Mn adalah nilai metrik pada kondisi normal, dan Mc adalah nilai metrik pada kondisi gangguan visual tertentu. Evaluasi robustness menggunakan lima kondisi, yaitu normal, dimmed 50%, brightened 50%, low contrast, dan Gaussian noise.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Evaluasi pada Kondisi Normal

Evaluasi pada test set normal menunjukkan bahwa setiap model memiliki karakteristik berbeda. YOLO11n memperoleh precision 0,4386, recall 0,4466, F1-score 0,4426, mAP@50 0,4301, dan mAP@50:95 0,3135. YOLO11s meningkatkan precision menjadi 0,5047 dan F1-score menjadi 0,4777, dengan mAP@50 sebesar 0,4467 dan mAP@50:95 sebesar 0,3369. RT-DETR-l memperoleh precision tertinggi 0,5329, mAP@50 tertinggi 0,4484, dan mAP@50:95 tertinggi 0,3576. Hasil ini menunjukkan bahwa RT-DETR-l unggul pada kualitas deteksi secara keseluruhan, terutama pada metrik localization yang lebih ketat.

Performa YOLO11s yang lebih baik daripada YOLO11n mengindikasikan bahwa peningkatan kapasitas model membantu memperbaiki kualitas prediksi. Sementara itu, RT-DETR-l menghasilkan mAP@50:95 tertinggi karena pendekatan transformer mampu memodelkan relasi global antar area citra. Meskipun demikian, recall RT-DETR-l lebih rendah dibandingkan YOLO11n dan YOLO11s, sehingga model ini masih berpotensi melewati sebagian objek pada kondisi tertentu. Hasil evaluasi normal ditampilkan pada Tabel 3.

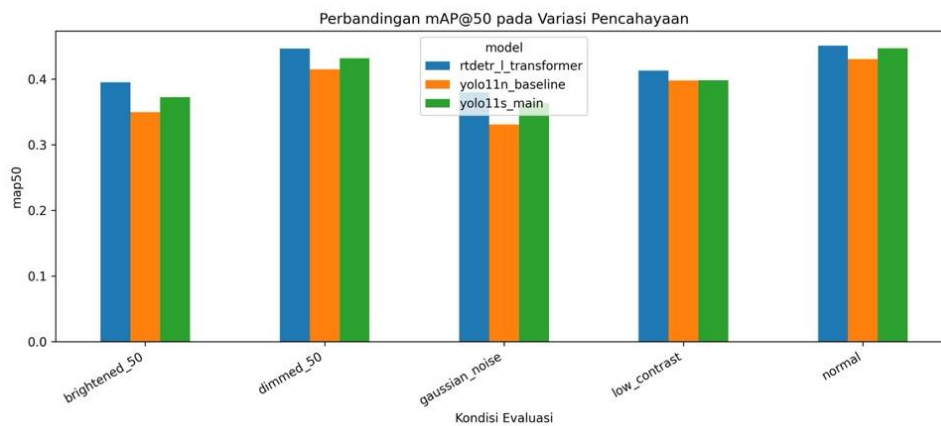
Tabel 3. Hasil evaluasi model pada test set normal

Model	Precision	Recall	F1-score	mAP@50	mAP@50:95
YOLO11n	0,4386	0,4466	0,4426	0,4301	0,3135
YOLO11s	0,5047	0,4534	0,4777	0,4467	0,3369
RT-DETR-l	0,5329	0,4190	0,4691	0,4484	0,3576

3.2 Evaluasi Robustness terhadap Variasi Pencahayaan

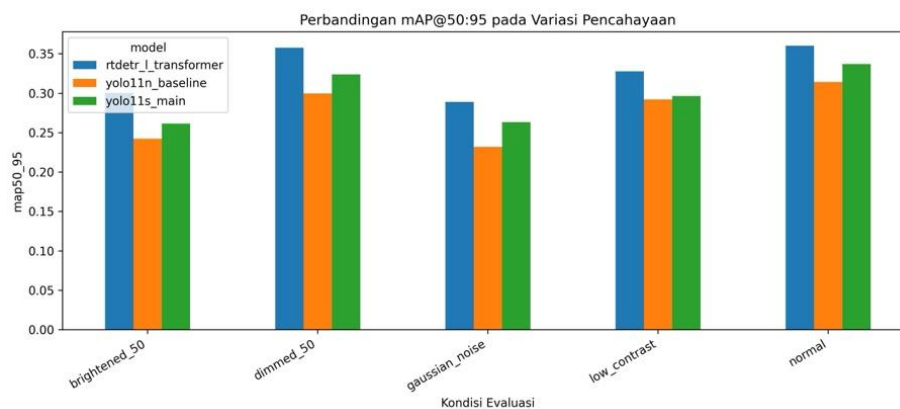
Pengujian robustness dilakukan untuk mengetahui stabilitas model ketika kualitas citra mengalami perubahan. Pada lingkungan nyata, citra sampah dapat diperoleh pada pencahayaan redup, pencahayaan berlebih, kontras rendah, atau kondisi kamera yang menghasilkan noise. Oleh karena itu, nilai mAP pada data normal perlu dilengkapi dengan analisis robustness agar pemilihan model lebih representatif.

Gambar 3 menunjukkan perbandingan mAP@50 pada lima kondisi evaluasi. RT-DETR-l menghasilkan mAP@50 tertinggi pada seluruh skenario. Pada kondisi normal, RT-DETR-l mencapai 0,4507, diikuti YOLO11s sebesar 0,4469 dan YOLO11n sebesar 0,4302. Pada kondisi dimmed 50%, RT-DETR-l tetap stabil dengan mAP@50 0,4463, sedangkan YOLO11s memperoleh 0,4319 dan YOLO11n memperoleh 0,4147. Kondisi brightened 50% menyebabkan penurunan pada seluruh model dengan nilai 0,3950 untuk RT-DETR-l, 0,3723 untuk YOLO11s, dan 0,3497 untuk YOLO11n. Pada kondisi low contrast, RT-DETR-l memperoleh 0,4129, sedangkan YOLO11s dan YOLO11n masing-masing memperoleh 0,3981 dan 0,3979. Gaussian noise menjadi skenario yang paling menantang karena nilai mAP@50 turun menjadi 0,3798 pada RT-DETR-l, 0,3628 pada YOLO11s, dan 0,3308 pada YOLO11n.



Gambar 3. Perbandingan mAP@50 pada variasi pencahayaan

Gambar 4 memperlihatkan nilai mAP@50:95 yang lebih ketat karena memperhitungkan rentang IoU 0,50 sampai 0,95. Pola performa tetap konsisten, yaitu RT-DETR-l unggul pada semua kondisi. Pada kondisi normal, RT-DETR-l memperoleh mAP@50:95 sebesar 0,3599, sedangkan YOLO11s memperoleh 0,3369 dan YOLO11n memperoleh 0,3140. Kondisi dimmed 50% hanya menurunkan RT-DETR-l sedikit menjadi 0,3575. Sebaliknya, brightened 50% dan Gaussian noise menurunkan performa lebih besar karena keduanya dapat merusak detail tepi, tekstur, dan kualitas lokalisasi objek.



Gambar 4. Perbandingan mAP@50:95 pada variasi pencahayaan

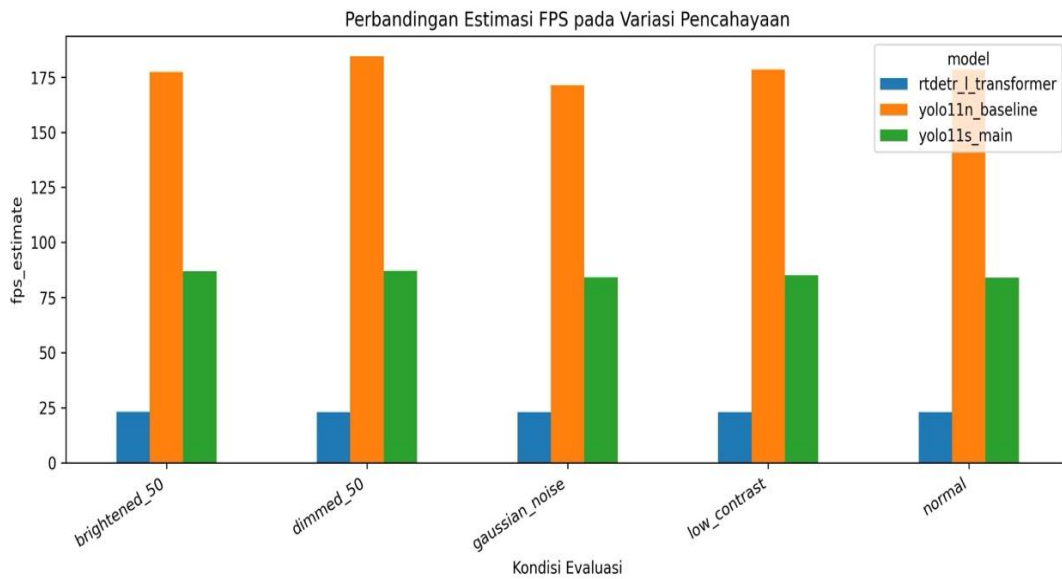
Tabel 4. Hasil evaluasi robustness berdasarkan mAP@50

Kondisi	YOLO11n	YOLO11s	RT-DETR-l
Normal	0,4302	0,4469	0,4507
Dimmed 50%	0,4147	0,4319	0,4463
Brightened 50%	0,3497	0,3723	0,3950
Low Contrast	0,3979	0,3981	0,4129
Gaussian Noise	0,3308	0,3628	0,3798

3.3 Analisis Efisiensi Inferensi

Kecepatan inferensi menjadi faktor penting dalam sistem real-time. Gambar 5 menunjukkan bahwa YOLO11n memperoleh FPS tertinggi pada seluruh kondisi evaluasi. Pada kondisi normal, YOLO11n mencapai 178,32 FPS, YOLO11s mencapai 84,05 FPS, sedangkan RT-DETR-l mencapai 23,01 FPS. Perbedaan ini menunjukkan bahwa model dengan kapasitas lebih besar dan mekanisme transformer memiliki konsekuensi terhadap kecepatan proses inferensi.

YOLO11n sesuai untuk perangkat dengan sumber daya komputasi terbatas karena mampu memproses citra dengan sangat cepat. YOLO11s lebih sesuai untuk skenario yang membutuhkan keseimbangan antara akurasi dan kecepatan. RT-DETR-l lebih tepat untuk kebutuhan yang mengutamakan akurasi dan stabilitas terhadap variasi visual, tetapi kurang ideal untuk perangkat edge yang terbatas. Hasil FPS pada setiap kondisi ditampilkan pada Tabel 5.



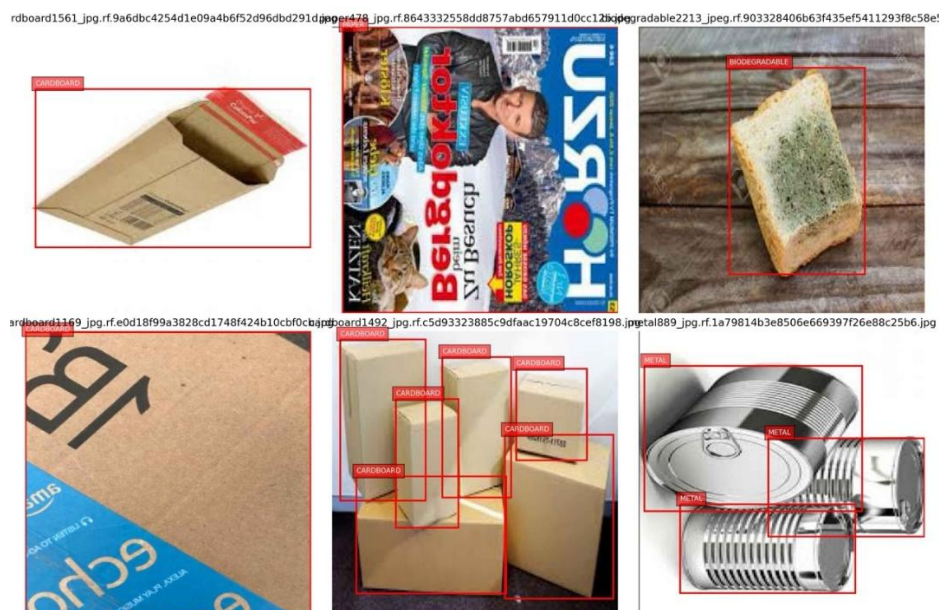
Gambar 5. Perbandingan estimasi FPS pada variasi pencahayaan

Tabel 5. Perbandingan FPS pada kondisi evaluasi

Kondisi	YOLO11n	YOLO11s	RT-DETR-1
Normal	178,32	84,05	23,01
Dimmed 50%	184,33	87,15	22,99
Brightened 50%	177,30	86,90	23,13
Low Contrast	178,40	85,18	23,09
Gaussian Noise	171,35	84,27	22,99

3.4 Visualisasi Prediksi Sampel

Visualisasi prediksi digunakan untuk memeriksa secara kualitatif kemampuan model dalam mengenali objek. Gambar 6 memperlihatkan contoh citra yang berisi objek cardboard, paper, biodegradable, dan metal. Beberapa citra memiliki lebih dari satu bounding box, sehingga pendekatan object detection lebih tepat dibandingkan klasifikasi citra tunggal. Visualisasi juga menunjukkan adanya tantangan berupa objek bertumpuk, latar belakang yang tidak seragam, dan variasi tekstur. Kondisi tersebut menjelaskan mengapa evaluasi tidak cukup hanya menggunakan satu metrik, tetapi perlu dilengkapi dengan analisis per kelas, robustness, dan kecepatan inferensi.



Gambar 6. Contoh visualisasi prediksi objek sampah

3.5 Pembahasan Komparatif

Berdasarkan seluruh hasil, RT-DETR-1 menjadi model paling unggul dari sisi akurasi dan robustness. Keunggulan tersebut terlihat dari nilai mAP@50 dan mAP@50:95 yang konsisten lebih tinggi pada kondisi normal maupun variasi pencahayaan. Hal ini dapat dikaitkan dengan kemampuan transformer dalam memodelkan hubungan global antar fitur visual. Pada citra sampah yang memiliki bentuk dan latar belakang bervariasi, konteks global membantu model membedakan objek yang memiliki tekstur mirip atau saling berdekatan.

Namun, RT-DETR-1 memiliki kekurangan pada efisiensi inferensi. FPS RT-DETR-1 berada pada kisaran 23 FPS, jauh lebih rendah dibandingkan YOLO11n yang mencapai lebih dari 170 FPS dan YOLO11s yang mencapai lebih dari 80 FPS. Dengan demikian, RT-DETR-1 lebih sesuai untuk sistem yang memprioritaskan akurasi, sedangkan YOLO11n lebih sesuai untuk sistem yang membutuhkan pemrosesan cepat. YOLO11s dapat diposisikan sebagai pilihan kompromi karena memberikan akurasi lebih baik daripada YOLO11n dengan kecepatan yang masih tinggi.

Dari sisi robustness, kondisi brightened 50% dan Gaussian noise menyebabkan penurunan performa paling besar. Pencahayaan berlebih dapat menghilangkan detail tepi objek dan menurunkan kontras lokal, sedangkan noise mengganggu tekstur yang dibutuhkan model untuk membedakan kelas. Kondisi dimmed 50% relatif lebih mudah ditangani oleh RT-DETR-1, terlihat dari penurunan mAP yang sangat kecil. Temuan ini memperlihatkan bahwa kualitas pencahayaan tidak hanya berpengaruh terhadap confidence prediksi, tetapi juga terhadap kualitas lokalisasi bounding box.

Faktor class imbalance juga perlu diperhatikan. Kelas BIODEGRADABLE memiliki jumlah bounding box jauh lebih besar dibandingkan kelas lain. Ketidakseimbangan tersebut dapat membuat model lebih sering mempelajari pola kelas mayoritas dan kurang optimal pada kelas minoritas. Masalah imbalance dalam object detection dapat muncul pada tingkat kelas, jumlah objek, ukuran objek, maupun kualitas anotasi [14]. Oleh karena itu, penelitian lanjutan dapat menggunakan strategi class-aware augmentation, re-sampling, pengayaan data kelas minoritas, atau loss function yang lebih sensitif terhadap ketidakseimbangan.

Secara praktis, hasil penelitian ini menunjukkan bahwa pemilihan model tidak dapat hanya didasarkan pada nilai akurasi tertinggi. RT-DETR-1 direkomendasikan ketika sistem membutuhkan stabilitas dan akurasi yang lebih baik, YOLO11n direkomendasikan untuk perangkat real-time ringan, sedangkan YOLO11s direkomendasikan sebagai model tengah yang menyeimbangkan akurasi dan efisiensi. Analisis trade-off tersebut menjadi kontribusi utama penelitian karena memberikan dasar yang lebih realistis untuk implementasi sistem deteksi sampah otomatis.

4. KESIMPULAN

Penelitian ini telah mengevaluasi performa dan robustness YOLO11n, YOLO11s, dan RT-DETR-1 pada deteksi sampah real-time dengan enam kelas objek. Hasil pengujian normal menunjukkan bahwa RT-DETR-1 memperoleh performa akurasi terbaik dengan precision 0,5329, mAP@50 0,4484, dan mAP@50:95 0,3576. Pada evaluasi robustness, RT-DETR-1 juga menunjukkan stabilitas paling baik pada variasi pencahayaan dan noise, khususnya pada kondisi dimmed 50% dengan mAP@50 0,4463 dan mAP@50:95 0,3575. Walaupun demikian, RT-DETR-1 memiliki kelemahan pada kecepatan inferensi karena hanya mencapai sekitar 23 FPS. Sebaliknya, YOLO11n menjadi model paling efisien dengan 178,32 FPS pada kondisi normal, tetapi akurasinya lebih rendah dibandingkan dua model lainnya. YOLO11s berada pada posisi kompromi karena memiliki performa lebih baik daripada YOLO11n dan kecepatan yang jauh lebih tinggi daripada RT-DETR-1. Dengan demikian, pemilihan model perlu disesuaikan dengan kebutuhan implementasi: RT-DETR-1 untuk akurasi dan robustness, YOLO11n untuk sistem real-time ringan, serta YOLO11s untuk keseimbangan antara akurasi dan efisiensi. Penelitian selanjutnya disarankan untuk menangani class imbalance melalui augmentasi berbasis kelas, menambah data dari kondisi lingkungan nyata, dan menguji model pada perangkat edge agar hasil implementasi lebih komprehensif.

REFERENCES

- [1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
- [2] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020, doi: 10.48550/arXiv.2004.10934.

- [3] 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 (CVPR), 2023, pp. 7464-7475, doi: 10.1109/CVPR52729.2023.00721.
- [4] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," in Computer Vision - ECCV 2024, 2024, pp. 1-21, doi: 10.1007/978-3-031-72751-1_1.
- [5] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "YOLOv10: Real-Time End-to-End Object Detection," in Advances in Neural Information Processing Systems, 2024, doi: 10.48550/arXiv.2405.14458.
- [6] Ultralytics, "Ultralytics YOLO11 Documentation," 2024. [Online]. Available: <https://docs.ultralytics.com/models/yolo11>. Accessed: Jul. 3, 2026.
- [7] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers," in Computer Vision - ECCV 2020, 2020, pp. 213-229, doi: 10.1007/978-3-030-58452-8_13.
- [8] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs Beat YOLOs on Real-Time Object Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 16965-16974, doi: 10.1109/CVPR52733.2024.01605.
- [9] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context," in Computer Vision - ECCV 2014, 2014, pp. 740-755, doi: 10.1007/978-3-319-10602-1_48.
- [10] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," International Journal of Computer Vision, vol. 88, pp. 303-338, 2010, doi: 10.1007/s11263-009-0275-4.
- [11] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A Survey on Performance Metrics for Object-Detection Algorithms," Electronics, vol. 10, no. 3, p. 279, 2021, doi: 10.3390/electronics10030279.
- [12] C. Shorten and T. M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," Journal of Big Data, vol. 6, no. 60, 2019, doi: 10.1186/s40537-019-0197-0.
- [13] K. Oksuz, B. C. Cam, S. Kalkan, and E. Akbas, "Imbalance Problems in Object Detection: A Review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 10, pp. 3388-3415, 2021, doi: 10.1109/TPAMI.2020.2981890.
- [14] A. Rahmatulloh, I. Darmawan, A. P. Aldya, and F. M. S. Nursuwars, "WasteInNet: Deep Learning Model for Real-Time Identification of Various Types of Waste," Cleaner Waste Systems, vol. 10, p. 100198, 2025, doi: 10.1016/j.clw.2024.100198.
- [15] R. G. Girshick, "Fast R-CNN," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.