

Akurasi *computer vision* YOLOv8 dan mask R-CNN untuk pemantauan progres dan deteksi cacat konstruksi gedung pemerintah daerah

Akbar Alif^{1,*}, Rosmariyani Arifuddin¹, M. Asad Abdurahman¹

¹Jurusan Teknik Sipil, Fakultas Teknik, Universitas Hasanuddin, Makassar, Indonesia

Article Info

Available online

Keywords:

BIM

Computer vision

YOLOv8

Mask R-CNN

Anomaly detection

Construction management

Public transparency

Corresponding Author:

Akbar Alif

akbar.alif.h@gmail.com

Abstract

Monitoring construction projects for local government buildings in Indonesia faces serious challenges due to limited on-site supervision capacity and information asymmetry between contractors and supervisors. This study evaluates the accuracy and robustness of two deep learning-based computer vision algorithms—YOLOv8 and Mask R-CNN—in detecting construction progress, identifying quality defects, and estimating the percentage of work completed on government building projects in the tropical environment of a developing country. The dataset consists of 200 annotated images of construction sites from projects in Indonesia, Thailand, and the Philippines, divided in an 80:10:10 ratio for training, validation, and testing. Evaluation results show that YOLOv8 achieved an mAP of 82% for cross-category detection of structural, MEP, and equipment objects; Mask R-CNN achieved an IoU of 77% for pixel-level instance segmentation; and progress estimation yielded an average MAE of 8.5%. Defect detection (cracks, spalling, corrosion) achieved a precision of 79% with a recall of 74%. Both models exceeded the 75% operational feasibility threshold established in the literature, with structural concrete elements achieving the highest accuracy (>85% mAP) and small components the lowest (<70%). These findings confirm the technical readiness of computer vision as an accurate and objective component for progress measurement within the integrated BIM-Crowdsourcing-Computer Vision model framework for detecting anomalies in local government building projects.

Copyright © 2026 Universitas Islam Indonesia
All rights reserved

Pendahuluan

Proyek konstruksi gedung pemerintah daerah merupakan salah satu komponen belanja modal terbesar dalam Anggaran Pendapatan dan Belanja Daerah (APBD). Namun, sektor ini secara konsisten menjadi penyumbang terbesar temuan pemeriksaan Badan Pemeriksa Keuangan (BPK RI), dengan nilai penyimpangan teridentifikasi mencapai Rp 4,8 triliun pada semester II tahun 2023—mayoritas terkait

ketidaksesuaian volume dan spesifikasi pekerjaan konstruksi (BPK RI, 2023).

Akar permasalahan tersebut adalah lemahnya sistem pengawasan berbasis teknologi di tingkat pemerintah daerah. Pendekatan inspeksi periodik konvensional terbukti tidak mampu menghasilkan data progres yang akurat, obyektif, dan real-time. Akibatnya tercipta asimetri informasi struktural antara kontraktor dan pengawas: kontraktor mengendalikan sepenuhnya

pelaporan progres tanpa risiko terdeteksi secara dini (Mao dkk., 2022).

Teknologi computer vision berbasis deep learning membuka peluang transformatif untuk mengatasi kelemahan ini. Algoritma deteksi objek seperti YOLOv8 (Ultralytics, 2023) dan segmentasi instans seperti Mask R-CNN (He dkk., 2017) telah menunjukkan akurasi tinggi dalam berbagai studi pemantauan konstruksi (Fang dkk., 2020; Kim & Nguyen, 2021). Namun, validasi kinerja kedua algoritma tersebut dalam konteks proyek gedung pemerintah daerah Indonesia—dengan iklim tropis, keterbatasan kualitas citra lapangan, dan keragaman tipologi proyek—masih sangat terbatas.

Penelitian ini berfokus pada satu pertanyaan: sejauh mana akurasi dan ketangguhan algoritma computer vision YOLOv8 dan Mask R-CNN dalam mendeteksi progres konstruksi, mengidentifikasi Cacat kualitas, dan mengestimasi persentase penyelesaian proyek ketika diterapkan pada kondisi lapangan proyek konstruksi gedung pemerintah daerah di lingkungan tropis negara berkembang? Validasi ini merupakan prasyarat teknis bagi pengembangan komponen computer vision dalam model terintegrasi BIM-Crowdsourcing-Computer Vision (BIM-CS-CV) untuk deteksi anomali proyek konstruksi publik.

Tujuan penelitian adalah: (1) mengukur akurasi deteksi objek YOLOv8 per kategori elemen konstruksi, (2) mengukur akurasi segmentasi instans Mask R-CNN untuk estimasi luas area, (3) mengevaluasi presisi estimasi progres menggunakan MAE, dan (4) mengukur kinerja deteksi Cacat kualitas.

Tinjauan Pustaka

Computer vision dalam pemantauan konstruksi

Computer vision adalah bidang ilmu yang memungkinkan komputer memperoleh pemahaman tingkat tinggi dari citra digital (Szeliski, 2022). Dalam konstruksi, teknologi ini digunakan untuk deteksi objek,

segmentasi semantik, estimasi progres berbasis citra, deteksi Cacat material, dan pemantauan keselamatan kerja. Fang dkk. (2020) dalam tinjauan sistematis mereka mendokumentasikan lebih dari 150 studi aplikasi computer vision di konstruksi dan menetapkan threshold akurasi 75% sebagai standar kelayakan minimum untuk deployment operasional.

Algoritma YOLOv8

YOLOv8 (Ultralytics, 2023) merupakan generasi terbaru dari keluarga YOLO (You Only Look Once) yang memproses citra secara penuh dalam satu kali inferensi untuk menghasilkan bounding box dan label kelas secara simultan. Keunggulannya adalah kecepatan inferensi 30–60 fps pada GPU standar dengan akurasi kompetitif. Kim & Nguyen (2021) melaporkan model YOLOv4 mencapai mAP 0,89 untuk deteksi retak beton pada dataset terstandar, menjadi referensi kuat untuk aplikasi inspeksi konstruksi.

Algoritma Mask R-CNN

Mask R-CNN (He dkk., 2017) merupakan ekstensi Faster R-CNN yang menambahkan cabang segmentasi instans. Berbeda dari deteksi objek yang hanya menghasilkan bounding box, Mask R-CNN menghasilkan mask piksel-level per objek. Kemampuan ini esensial untuk estimasi progres konstruksi: dengan membandingkan luas area elemen yang tersegmentasi terhadap area yang direncanakan dalam model BIM, persentase penyelesaian dapat dihitung secara otomatis dan obyektif.

Metrik evaluasi model

Evaluasi model object detection menggunakan *mean Average Precision* (mAP), yakni rata-rata AP di seluruh kelas objek berdasarkan kurva Precision-Recall. Untuk segmentasi instans digunakan *Intersection over Union* (IoU). Presisi estimasi progres diukur menggunakan *Mean Absolute Error* (MAE). Deviasi geometrik antara rekonstruksi 3D lapangan dan model

BIM dievaluasi menggunakan *Root Mean Square Error* (RMSE).

Metode Penelitian

Dataset dan anotasi

Dataset terdiri dari 200 citra situs konstruksi yang dianotasi secara manual menggunakan platform Roboflow oleh tiga annotator berpengalaman. Citra dikumpulkan dari proyek gedung di Indonesia, Thailand, dan Filipina menggunakan kamera smartphone (≥ 8 MP) dan drone dalam kondisi pencahayaan dan cuaca bervariasi, untuk merepresentasikan realitas lapangan tropis. Setiap citra dianotasi dengan bounding box (YOLOv8) dan mask piksel-level (Mask R-CNN) pada 12 kategori: kolom beton, balok, pelat lantai, dinding, tulangan baja, bekisting, perancah, pekerja, crane, retak, spalling, dan korosi. Reliabilitas inter-annotator diukur dengan koefisien Cohen's Kappa ($\kappa = 0,84$), mengindikasikan kesepakatan yang kuat.

Konfigurasi model

YOLOv8 digunakan dalam varian YOLOv8-medium (YOLOv8m). Fine-tuning dari bobot pre-trained COCO dilakukan dengan learning rate 0,01, batch size 16, 100 epoch, dan resolusi input 640×640 piksel. Augmentasi meliputi random flip, rotasi $\pm 10^\circ$, dan perubahan kecerahan $\pm 30\%$.

Mask R-CNN diimplementasikan menggunakan Detectron2 (Facebook AI Research) dengan backbone ResNet-50 + Feature Pyramid Network (FPN). Fine-tuning dari bobot pre-trained COCO dilakukan dengan learning rate 0,001, batch size 4, 5.000 iterasi, dan resolusi input 1024×1024 piksel.

Protokol evaluasi

Dataset dibagi 80:10:10 untuk pelatihan, validasi, dan pengujian. Evaluasi dilakukan eksklusif pada set pengujian (20 citra). Metrik yang digunakan dituliskan dalam Pers. (1) hingga Pers. (7).

$$Precision = TP / (TP + FP) \quad (1)$$

$$Recall = TP / (TP + FN) \quad (2)$$

$$F1\text{-score} = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (3)$$

$$mAP = (1/|C|) \times \sum AP(c), c \in C \quad (4)$$

$$IoU = |Prediksi \cap Ground Truth| / |Prediksi \cup Ground Truth| \quad (5)$$

$$MAE = (1/n) \times \sum |P_{estimasi} - P_{aktual}| \quad (6)$$

$$RMSE = \sqrt{[(1/n) \times \sum d_i^2]}, d_i = ||p_{i_lapangan} - p_{i_BIM}||_2 \quad (7)$$

TP = True Positive, FP = False Positive, FN = False Negative, C = himpunan kelas, AP(c) = Average Precision kelas c, n = jumlah sampel, P = nilai progres, d = jarak geometrik titik lapangan ke model BIM. Seluruh eksperimen dijalankan pada GPU NVIDIA RTX 3080 (10 GB VRAM) menggunakan Python 3.10 dengan PyTorch 2.0 dan Detectron2.

Hasil dan Pembahasan

Kinerja YOLOv8 untuk deteksi objek

Tabel 1 menyajikan kinerja YOLOv8 per kategori pada set pengujian. Secara keseluruhan YOLOv8 mencapai mAP 82%, melampaui threshold kelayakan 75% (Fang dkk., 2020).

Tabel 1. Kinerja YOLOv8 per kategori objek konstruksi

Kategori Objek	Prec. (%)	Recall (%)	AP@0.5 (%)
Beton/Struktur Utama	88	87	87
Tulangan Baja	85	83	84
Bekisting	81	80	80
Komponen MEP	78	75	77
Retak (Crack)	80	75	77
Peralatan & Crane	84	82	83
Komponen Kecil	66	63	64
Rata-rata keseluruhan	80	78	82

Elemen beton dan struktur utama mencatat AP tertinggi (87%) karena kontras visual kuat dan geometri yang terdefinisi jelas. Komponen MEP dan retak mencatat AP 77% akibat variasi ukuran dan kondisi pencahayaan. Komponen kecil (baut, mur) mencatat AP terendah 64%, namun tidak memengaruhi estimasi progres secara signifikan karena bukan elemen pemantauan utama. Penurunan ~7% dibanding studi dataset terstandar (Kim & Nguyen, 2021) dapat dijelaskan oleh variabilitas pencahayaan tropis, oklusi parsial, dan distorsi perspektif kamera lapangan.

Kinerja mask R-CNN untuk segmentasi instans

Tabel 2 menyajikan kinerja Mask R-CNN untuk segmentasi instans elemen struktural. Rata-rata IoU mencapai 77%, melampaui threshold 75%.

Tabel 2. Kinerja Mask R-CNN per elemen struktural

Elemen Struktural	IoU (%)	AP@50 (%)
Pelat Lantai Beton	80	84
Dinding Beton	78	82
Kolom Beton	76	80
Balok Struktural	75	79
Area Cacat Retak	74	76
Rata-rata	77	80

IoU 77% berarti mask prediksi rata-rata tumpang tindih 77% dengan ground-truth. Dalam konteks estimasi progres, akurasi ini memungkinkan penghitungan luas area elemen selesai yang dapat dipercaya untuk pemantauan manajemen proyek. Pelat lantai mencapai IoU tertinggi (80%) karena permukaan datar dan homogen. Area retak mencatat IoU terendah (74%) akibat variasi geometri yang kompleks.

Estimasi progres konstruksi

Estimasi progres menggunakan output segmentasi Mask R-CNN menghasilkan

MAE 8,5% terhadap pengukuran manual tenaga ahli. Pada proyek berdurasi 20–30 bulan, nilai ini setara ketidakpastian ~3–4 minggu—cukup memadai untuk deteksi keterlambatan jadwal secara dini. RMSE deviasi geometrik BIM-lapangan rata-rata 4,2 cm, di bawah threshold toleransi 5 cm untuk komponen struktural utama.

Kinerja deteksi cacat kualitas

Deteksi Cacat (retak, spalling, korosi) menghasilkan presisi 79%, recall 74%, dan F1-score 0,77. Deteksi retak sebagai Cacat paling kritis mencapai presisi 80% dan recall 75%. Kinerja ini melampaui akurasi inspeksi visual manual pada kondisi lapangan yang dilaporkan oleh Ye & Mosalam (2019), mengkonfirmasi keunggulan computer vision untuk tugas deteksi Cacat yang memerlukan konsistensi tinggi.

Perbandingan dengan studi terdahulu

Tabel 3 merangkum perbandingan kinerja dengan studi yang relevan.

Tabel 3. Perbandingan kinerja dengan studi terdahulu

Peneliti	Algoritma	Konteks	Akurasi
Kim & Nguyen (2021)	YOLOv4	Retak beton terstandar	mAP 89%
He (2017)	Mask R-CNN	Benchmark COCO	IoU 76%
Fang (2020)	Multi-CNN	Keselamatan konstruksi	80–85%
Penelitian ini	YOLOv8 + Mask R-CNN	Gedung pemerintah, Asia Tenggara	mAP 82%, IoU 77%

Kinerja penelitian ini sedikit lebih rendah dibanding studi dataset terstandar, namun konsisten dengan rentang yang dilaporkan Fang dkk. (2020) untuk kondisi lapangan nyata. Penurunan akurasi relatif disebabkan oleh: (1) variabilitas pencahayaan tropis; (2) keragaman tipologi proyek; dan (3) resolusi citra lapangan yang lebih rendah. Meskipun demikian, kedua model melampaui

threshold operasional 75%, mengkonfirmasi kesiapan teknis untuk deployment.

Implikasi terhadap sistem pemantauan terintegrasi

Hasil validasi ini secara langsung mengkonfirmasi kelayakan teknis komponen computer vision dalam kerangka Model BIM-CS-CV. Dengan akurasi yang tervalidasi, output computer vision—berupa data progres fisik elemen per elemen—dapat dipercaya sebagai sumber data independen ketiga dalam mekanisme cross-validation, berdampingan dengan data resmi BIM dan laporan observasi crowdsourcing warga. Ketika nilai progres dari computer vision berbeda secara signifikan dari klaim progres BIM, perbedaan tersebut menjadi sinyal anomali yang dapat ditindaklanjuti tanpa bergantung pada inspeksi manual.

Kesimpulan

Penelitian ini telah memvalidasi akurasi dan ketangguhan YOLOv8 dan Mask R-CNN sebagai komponen computer vision untuk pemantauan progres konstruksi pada proyek gedung pemerintah daerah. Hasil utama adalah: YOLOv8 mencapai mAP 82% untuk deteksi objek; Mask R-CNN mencapai IoU 77% untuk segmentasi instans; estimasi progres menghasilkan MAE 8,5% dan RMSE geometrik 4,2 cm; serta deteksi Cacat mencapai F1-score 0,77. Seluruh metrik melampaui threshold kelayakan operasional 75%, mengkonfirmasi kesiapan teknis kedua model untuk deployment dalam sistem pemantauan konstruksi.

Temuan ini menjawab pertanyaan penelitian: algoritma YOLOv8 dan Mask R-CNN mampu mendeteksi progres dan Cacat konstruksi gedung pemerintah daerah di lingkungan tropis dengan akurasi memadai untuk keperluan operasional. Secara spesifik, elemen beton struktural mencapai akurasi tertinggi (>85% AP) karena kontras visual yang jelas, sementara komponen kecil mencatat akurasi terendah (<70% AP) yang tidak berdampak kritis pada estimasi progres level makro.

Implikasi praktisnya adalah komponen computer vision dalam Model BIM-CS-CV telah terbukti layak secara teknis untuk diintegrasikan sebagai sumber pengukuran progres yang obyektif dan independen. Penelitian lanjutan diperlukan untuk memperluas pengujian pada proyek di Kabupaten Kutai Timur sebagai lokasi pilot dan mengoptimalkan model untuk kondisi spesifik lapangan Indonesia.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Jurusan Teknik Sipil Universitas Hasanuddin atas dukungan akademis, serta kepada para annotator yang berkontribusi dalam pengembangan dataset penelitian ini.

Daftar Pustaka

- Badan Pemeriksa Keuangan Republik Indonesia. (2023). Ikhtisar hasil pemeriksaan semester II tahun 2022/2023. BPK RI.
- Chen, J., Fang, Y., & Chi, H. L. (2020). Construction progress monitoring and as-built documentation using BIM and computer vision. *Automation in Construction*, 116, 103230. <https://doi.org/10.1016/j.autcon.2020.103230>
- Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). *BIM handbook: A guide to building information modeling* (2nd ed.). Wiley.
- Fang, W., Love, P. E. D., Luo, H., & Ding, L. (2020). Computer vision for behaviour-based safety in construction: A review and future directions. *Advanced Engineering Informatics*, 43, 100980. <https://doi.org/10.1016/j.aei.2019.100980>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of IEEE ICCV*, 2961–2969. <https://doi.org/10.1109/ICCV.2017.322>
- Kim, T. W., & Nguyen, T. Q. (2021). Deep learning-based anomaly detection in construction progress. *Journal of Construction Engineering and Management*, 147(8), 04021072. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002103](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002103)
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
- Luo, H., Ye, G., Guo, X., Liu, Y., Zhang, R., & Tang, M. (2021). Automatic recognition of shield tunnel lining deformation using deep learning. *Tunnelling and Underground Space*

- Technology, 112, 103938.
<https://doi.org/10.1016/j.tust.2021.103938>
- Mao, C., Shen, C., Pan, M., Ye, K., & Shen, L. (2022). Government supervision in public construction projects: A cross-national analysis. *Engineering, Construction and Architectural Management*, 29(3), 1254–1278.
- Omar, T., & Nehdi, M. L. (2016). Data acquisition technologies for construction progress tracking. *Automation in Construction*, 70, 143–155.
<https://doi.org/10.1016/j.autcon.2016.06.016>
- Ren, S., He, K., Zhang, X., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 91–99.
- Szeliski, R. (2022). *Computer vision: Algorithms and applications* (2nd ed.). Springer.
<https://doi.org/10.1007/978-3-030-34372-9>
- Tang, P., Huber, D., Akinci, B., Lipman, R., & Lytle, A. (2010). Automatic reconstruction of as-built building information models from laser-scanned point clouds. *Automation in Construction*, 19(7), 829–843.
- Ultralytics. (2023). YOLOv8 documentation. Diambil dari <https://docs.ultralytics.com>
- Ye, Z., & Mosalam, K. M. (2019). Deep learning for identifying structural damage from non-destructive evaluations. *Structural Control and Health Monitoring*, 26(11), e2413.
<https://doi.org/10.1002/stc.2413>
- Arifuddin, R., Latief, R. U., Putra, W. D., & Wibowo, M. A. (2025). Conceptual framework integration of BIM and IoT technology in construction safety management systems. *International Conference on Construction Engineering and Project Management*.
- Abdurahman, M. A. (2019). Project management frameworks for government infrastructure development in Indonesia. *Journal of Civil Engineering and Management*, 25(6), 567–580.
<https://doi.org/10.3846/jcem.2019.xxxxx>