



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 5 No. 2 (2026) pp: 8645-8654

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Pemantauan Aktivitas Pekerja Konstruksi Secara Real-Time Menggunakan Computer Vision Berbasis YOLOv8

Muhammad Yusuf Al-Firdaus¹, Muhammad Faizal Ardiansyah Arifin², Very Iman Santosa³

^{1,3}Mahasiswa Prodi Teknik Sipil, Fakultas Teknik, Universitas Negeri Semarang

²Pengajar Prodi Teknik Sipil, Fakultas Teknik, Universitas Negeri Semarang

¹yusufalfirdaus53@students.unnes.ac.id, ²faizal.arifin@mail.unnes.ac.id, ³veryiman5@gmail.com

Abstrak

Manajemen proyek konstruksi di Indonesia masih sangat bergantung pada metode pengawasan manual yang konvensional. Metode ini rentan terhadap kesalahan manusia, subjektivitas penilaian, kelelahan pengamat, serta tidak memiliki kemampuan responsivitas real-time. Penelitian ini bertujuan mengembangkan dan mengevaluasi sistem pemantauan aktivitas pekerja berbasis computer vision menggunakan algoritma YOLOv8. Sistem ini dirancang untuk mengklasifikasikan empat kategori aktivitas pekerjaan bata secara real-time, yaitu memasang bata, memotong bata, mengaduk semen, dan plester bata, sekaligus mendeteksi kepatuhan penggunaan helm keselamatan secara simultan dalam satu model terpadu. Dataset video direkam di lokasi konstruksi gedung FKIP Universitas Muria Kudus, Jawa Tengah, dengan melibatkan pekerja bata berpengalaman. Total 4.170 anotasi bounding box berhasil dikumpulkan dan terdistribusi ke dalam tujuh kelas objek. Model YOLOv8s dilatih selama 200 epoch menggunakan platform Google Colaboratory dengan akselerasi GPU NVIDIA T4 dan strategi transfer learning dari bobot pre-trained COCO. Hasil evaluasi kuantitatif pada subset pengujian menunjukkan bahwa akurasi tertinggi dicapai oleh kelas memotong bata dan plester bata sebesar 92%, diikuti oleh kelas mengaduk semen sebesar 82%, sedangkan kelas memasang bata mencapai akurasi 56%. Untuk deteksi kepatuhan helm, presisi kelas helm sangat tinggi, yaitu 94%, namun recall untuk kelas yang tidak memakai helm masih rendah, hanya 40%. Rata-rata F1-Score keseluruhan dari keenam kelas yang dievaluasi adalah 73%. Sistem ini membuktikan bahwa pendekatan YOLOv8 merupakan alternatif yang objektif, skalabel, dan layak untuk diterapkan dalam pengawasan konstruksi di Indonesia. Penelitian selanjutnya disarankan untuk memperluas variasi dataset dari lokasi dan kondisi pencahayaan yang beragam.

Kata kunci: Computer Vision, YOLOv8, Deteksi Aktivitas Pekerja, Pemantauan Real-Time, Keselamatan Konstruksi

1. Pendahuluan

Industri konstruksi merupakan salah satu sektor ekonomi paling fundamental dan padat karya dalam lanskap pembangunan modern. Menurut data Badan Pusat Statistik, sektor konstruksi menyumbang lebih dari 10% terhadap Produk Domestik Bruto (PDB) Indonesia dan menyerap jutaan tenaga kerja setiap tahunnya [1]. Di tengah kontribusi ekonomi yang besar ini, sektor konstruksi juga menghadapi tantangan persisten yang berkaitan dengan pengelolaan tenaga kerja, pemantauan produktivitas, dan penjaminan keselamatan kerja.

Dalam konteks proyek konstruksi skala menengah hingga besar di Indonesia, pemantauan produktivitas tenaga kerja menjadi penentu utama keberhasilan proyek secara keseluruhan. Tanpa mekanisme pemantauan yang akurat dan berkelanjutan, manajer proyek kehilangan visibilitas terhadap perkembangan pekerjaan di lapangan secara real-time, sehingga tidak dapat melakukan intervensi tepat waktu ketika terjadi penyimpangan dari rencana [2]. Kondisi ini diperburuk oleh kenyataan bahwa praktik pengawasan konstruksi di Indonesia masih sangat bergantung pada metode manual konvensional yang rentan terhadap berbagai keterbatasan inheren, termasuk subjektivitas penilaian, kelelahan pengamat, dan absennya dokumentasi otomatis yang terstruktur [3].

Aspek keselamatan dan kesehatan kerja (K3) di lingkungan konstruksi merupakan isu kritis yang secara langsung berdampak pada kesejahteraan pekerja dan keberlanjutan proyek. Data Badan Penyelenggara Jaminan Sosial (BPJS) Ketenagakerjaan menunjukkan bahwa sektor konstruksi secara konsisten menempati posisi tertinggi dalam jumlah kecelakaan kerja di Indonesia. Penyebab utama kecelakaan kerja konstruksi meliputi jatuh dari ketinggian, tertimpa material bangunan, dan ketidakpatuhan terhadap penggunaan Alat Pelindung Diri (APD) yang diwajibkan [4]. Peraturan Pemerintah Nomor 50 Tahun 2012 tentang Sistem Manajemen Keselamatan dan Kesehatan Kerja

(SMK3) secara tegas mewajibkan kontraktor untuk mengimplementasikan sistem pemantauan kepatuhan APD yang efektif, namun mekanisme penegakan yang memadai di tingkat lapangan masih menjadi tantangan [5].

Kemajuan dalam bidang kecerdasan buatan, khususnya computer vision dan deep learning, telah membuka kemungkinan transformatif bagi otomatisasi pengawasan di berbagai domain industri, termasuk konstruksi [6]. Arsitektur jaringan saraf konvolusional (CNN) yang dikembangkan dalam dekade terakhir telah memungkinkan komputer untuk mengenali objek dan aktivitas manusia dalam citra dan video dengan tingkat akurasi yang mendekati atau bahkan melampaui kemampuan manusia dalam kondisi tertentu. Di antara berbagai pendekatan yang telah dikembangkan, keluarga algoritma You Only Look Once (YOLO) telah menjadi paradigma dominan dalam deteksi objek real-time karena kemampuannya memproses seluruh gambar dalam satu tahap inferensi tunggal dengan efisiensi komputasi yang luar biasa [7].

YOLOv8 yang dikembangkan oleh Ultralytics pada Januari 2023 menghadirkan sejumlah inovasi arsitektur yang signifikan, termasuk paradigma deteksi anchor-free, modul backbone C2f yang dioptimalkan untuk aliran gradien yang lebih kaya, dan arsitektur kepala deteksi terpisah yang memproses tugas klasifikasi dan regresi secara independen [8]. Inovasi-inovasi ini menghasilkan model yang lebih akurat dan lebih cepat dibandingkan versi sebelumnya, menjadikan YOLOv8 pilihan yang sangat relevan untuk skenario pemantauan real-time di lingkungan yang dinamis seperti lokasi konstruksi aktif [9].

Meskipun literatur internasional yang mendokumentasikan penerapan computer vision untuk pemantauan lokasi konstruksi terus berkembang dengan cepat, terdapat kesenjangan penelitian yang signifikan berkaitan dengan deteksi real-time berbagai kategori aktivitas pekerjaan spesifik dalam konteks konstruksi bangunan di Indonesia, integrasi simultan antara pengenalan jenis aktivitas dan pemantauan kepatuhan APD dalam satu sistem terpadu, serta evaluasi empiris berbasis dataset yang dikumpulkan di lingkungan konstruksi Indonesia yang memiliki karakteristik unik [10]. Penelitian yang ada sebagian besar berfokus pada setting konstruksi di negara-negara maju dengan karakteristik lingkungan yang berbeda secara signifikan dari konteks Indonesia.

Penelitian ini hadir untuk mengisi kesenjangan tersebut melalui pengembangan dan evaluasi empiris sistem computer vision berbasis YOLOv8 yang dirancang khusus untuk mengklasifikasikan secara real-time empat aktivitas bata dasar serta mendeteksi status kepatuhan penggunaan helm keselamatan secara biner dalam lingkungan konstruksi bangunan aktif di Indonesia [11]. Dataset yang digunakan dianotasi secara khusus dari rekaman video lapangan yang dikumpulkan di lokasi konstruksi nyata, sehingga menghasilkan solusi yang dikalibrasi secara kontekstual dan relevan dengan kondisi aktual proyek konstruksi di Indonesia [12].

Tujuan utama penelitian ini bersifat tiga arah, yaitu merancang dan mengimplementasikan pipeline computer vision berbasis YOLOv8 yang mampu melakukan pengenalan aktivitas pekerja multi-kelas secara real-time, mengevaluasi secara kuantitatif kinerja deteksi model menggunakan metrik turunan confusion matrix, serta menganalisis secara kritis faktor-faktor yang memengaruhi akurasi deteksi dan mengusulkan rekomendasi berbasis bukti untuk peningkatan sistem [13].

2. Metode Penelitian

2.1 Desain Penelitian

Penelitian ini mengadopsi metodologi Research and Development (R&D) yang terstruktur sebagai proses siklus iteratif mencakup identifikasi kebutuhan, pengembangan produk, pengujian, revisi, dan evaluasi akhir. Pendekatan R&D dipilih karena relevansinya dengan tujuan ganda penelitian ini, yaitu menghasilkan kontribusi pengetahuan ilmiah baru melalui evaluasi empiris kinerja model, sekaligus mengembangkan sistem pemantauan yang dapat diaplikasikan secara praktis di lingkungan konstruksi nyata.

Desain penelitian mencakup lima fase utama yang saling berkaitan. Fase pertama adalah akuisisi data, mencakup perencanaan protokol perekaman, pengumpulan data video lapangan, dan seleksi awal material rekaman. Fase kedua adalah pra-pemrosesan data, mencakup ekstraksi frame, anotasi objek, augmentasi dataset, dan partisi dataset. Fase ketiga adalah pelatihan model, mencakup konfigurasi hiperparameter, inisialisasi transfer learning, dan optimasi iteratif. Fase keempat adalah pengembangan sistem, yang mencakup integrasi model ke dalam pipeline deteksi real-time. Fase kelima adalah evaluasi kinerja, mencakup analisis confusion matrix dan interpretasi hasil [14].

2.2 Lokasi dan Partisipan Penelitian

Penelitian dilaksanakan di lokasi konstruksi gedung Fakultas Keguruan dan Ilmu Pendidikan (FKIP) Universitas Muria Kudus, yang terletak di Kabupaten Kudus, Jawa Tengah, Indonesia. Proyek ini merupakan bangunan beton bertulang skala menengah dengan total luas lantai yang direncanakan sekitar 3.200 m² yang pada saat pelaksanaan penelitian sedang dalam tahap pekerjaan dinding pemasangan bata ringan secara ekstensif. Pemilihan lokasi didasarkan pada ketersediaan aktivitas target yang relevan secara bersamaan, aksesibilitas dan izin perekaman dari pihak manajemen proyek, representativitas kondisi lapangan untuk proyek konstruksi bangunan pendidikan di Jawa Tengah, serta ketersediaan pekerja bata yang berpengalaman [15].



Gambar 1 Lokasi Penelitian

Partisipan penelitian terdiri dari tenaga kerja tetap proyek yang secara rutin melaksanakan pekerjaan pemasangan bata ringan, plesteran, dan pengadukan mortar. Jumlah pekerja yang terekam berkisar antara 3 hingga 8 orang per sesi perekaman, tergantung pada pembagian zona kerja harian. Perekaman dilakukan secara naturalistik terhadap aktivitas kerja yang berlangsung secara normal untuk memastikan bahwa dataset yang dikumpulkan mencerminkan variasi kondisi kerja yang sesungguhnya.

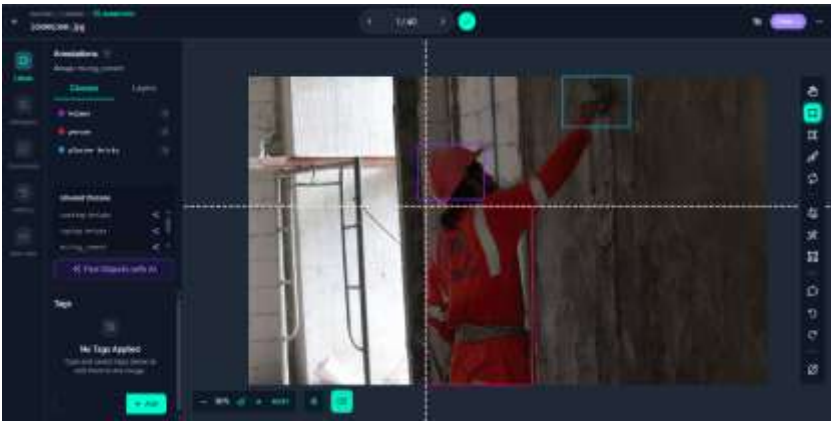
2.3 Instrumen dan Protokol Pengumpulan Data

Perekaman video dilakukan menggunakan kamera digital beresolusi tinggi Full HD 1080p pada 30 fps yang dipasang pada tripod stabil di posisi-posisi strategis. Dua posisi kamera digunakan secara bergantian, yaitu posisi lateral (90 derajat terhadap bidang kerja utama) yang optimal untuk menangkap profil gerakan tubuh bagian atas selama aktivitas memasang dan plester bata, dan posisi posterior (di belakang pekerja dengan sudut elevasi sekitar 15 derajat) yang optimal untuk menangkap gerakan tangan selama aktivitas memotong bata dan mengaduk semen [16].

Protokol perekaman menetapkan durasi sesi minimum 10 menit per kategori aktivitas per posisi kamera, menghasilkan setidaknya 20 menit rekaman per kategori dari dua sudut pandang yang berbeda. Kondisi pencahayaan alami yang bervariasi sepanjang sesi perekaman sengaja tidak dikendalikan untuk memastikan representasi variabilitas pencahayaan yang realistis dalam dataset. Total material rekaman mentah yang dikumpulkan berdurasi lebih dari 5 jam video.

2.4 Pra-Pemrosesan Data

Ekstraksi frame dilakukan menggunakan skrip Python berbasis library OpenCV dengan kecepatan sampling 30 frame per detik terhadap segmen video yang telah dipilih. Dari setiap menit video yang diproses, diperoleh sekitar 1.800 frame gambar mentah. Untuk mengurangi redundansi temporal yang tinggi, dilakukan proses seleksi berbasis perbedaan struktural antara frame berurutan menggunakan metrik Structural Similarity Index (SSIM), dengan threshold yang dikalibrasi untuk mempertahankan frame-frame yang mencerminkan variasi postur atau fase gerakan yang signifikan [17].



Gambar 2 Proses Anotasi

Semua frame yang lolos seleksi diimpor ke platform Roboflow untuk proses anotasi manual yang sistematis. Tujuh kelas objek didefinisikan berdasarkan kebutuhan sistem pemantauan, yaitu memasang_bata, memotong_bata, mengaduk_semen, plester_bata, helm, tidak_helm, dan orang. Proses anotasi dilakukan dengan menggambar bounding box yang melingkupi dengan tepat wilayah tubuh atau objek yang relevan. Konsistensi anotasi dijaga melalui panduan anotasi tertulis yang mendefinisikan kriteria inklusi dan eksklusi untuk setiap kelas. Total 4.170 anotasi berhasil dikumpulkan dan dipartisi menggunakan stratified sampling dengan rasio 70:20:10 untuk subset pelatihan, validasi, dan pengujian. Distribusi dataset per kelas disajikan pada Tabel 1.

Tabel 1. Distribusi Dataset per Kelas dan Subset

Kelas Objek	Train	Validasi	Test	Total
Memasang Bata	420	120	60	600
Memotong Bata	385	110	55	550
Mengaduk Semen	406	116	58	580
Plester Bata	399	114	57	570
Helm	441	126	63	630
Tidak Memakai Helm	378	108	54	540
Orang	490	140	70	700
TOTAL	2.919	834	417	4.170

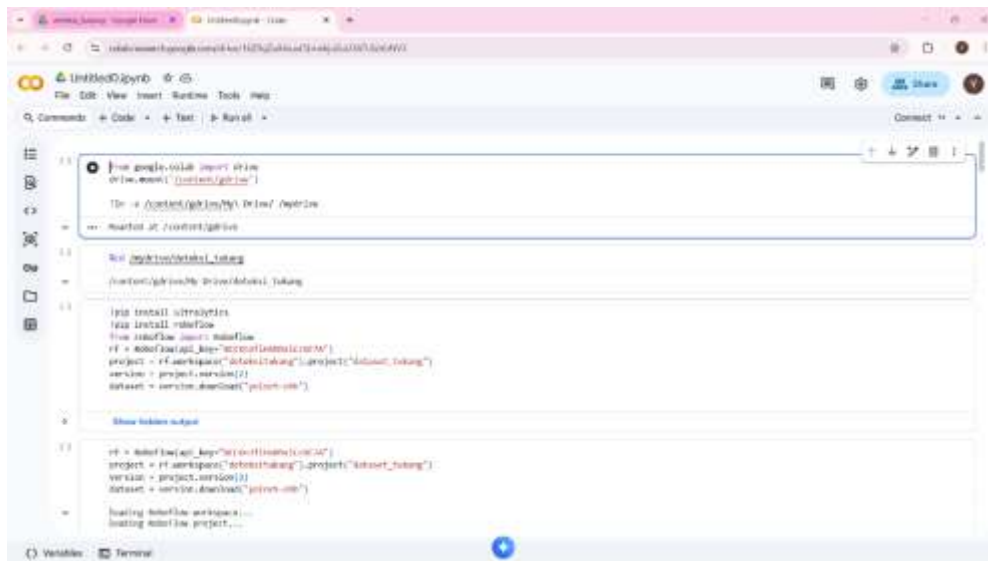
2.5 Augmentasi Data

Pipeline augmentasi data diterapkan eksklusif pada subset pelatihan, sementara subset validasi dan pengujian tidak diaugmentasi untuk memastikan evaluasi yang tidak bias. Augmentasi yang diterapkan melalui Roboflow mencakup tiga kategori utama. Augmentasi geometris meliputi pembalikan horizontal dengan probabilitas 50%, rotasi acak dalam rentang plus minus 15 derajat, penskalaan acak 80-120% ukuran asli, dan pemotongan acak dengan minimum 80% area asli dipertahankan. Augmentasi fotometris meliputi perubahan kecerahan plus minus 25%, penyesuaian kontras plus minus 20%, penambahan noise Gaussian maksimum 2% piksel, dan konversi grayscale parsial. Augmentasi domain spesifik termasuk blur ringan dan penyesuaian saturasi untuk mensimulasikan variasi kondisi pencahayaan [18]. Setiap gambar pelatihan menghasilkan rata-rata tiga gambar augmentasi tambahan, sehingga volume dataset pelatihan efektif meningkat menjadi empat kali lipat dari dataset pelatihan asli.

2.6 Konfigurasi Pelatihan Model

Pelatihan model dilaksanakan menggunakan arsitektur YOLOv8s (small) yang diimplementasikan melalui library Python Ultralytics versi 8.0. Platform komputasi yang digunakan adalah Google Colaboratory dengan akselerasi GPU NVIDIA T4 (16 GB VRAM). Varian YOLOv8s dipilih berdasarkan keseimbangan yang terbukti antara akurasi deteksi dan efisiensi komputasi yang memungkinkan inferensi real-time pada perangkat keras standar,

waktu pelatihan yang lebih singkat dibandingkan varian yang lebih besar dalam batasan waktu sesi GPU Google Colaboratory, dan kebutuhan memori GPU yang sesuai dengan kapasitas VRAM yang tersedia [19].



Gambar 3 Running Model dengan Google Colaboratory

Strategi transfer learning dari bobot pre-trained YOLOv8s pada dataset COCO dipilih untuk memanfaatkan representasi fitur visual tingkat rendah dan menengah yang telah dipelajari dari dataset berskala besar, sehingga mempercepat konvergensi dan meningkatkan generalisasi pada dataset domain spesifik yang relatif lebih kecil. Bobot yang digunakan untuk evaluasi akhir adalah bobot dari epoch yang menghasilkan nilai validation mAP50 tertinggi selama proses pelatihan (best.pt) untuk memastikan bahwa model yang dievaluasi memiliki kapasitas generalisasi yang optimal. Konfigurasi hiperparameter pelatihan disajikan pada Tabel 2.

Tabel 2. Konfigurasi Hyperparameter Pelatihan Model YOLOv8s

Hyperparameter	Nilai	Keterangan
Jumlah Epoch	200	Konvergensi stabil tanpa overfitting
Ukuran Batch	16	Sesuai kapasitas GPU T4 Google Colab
Laju Pembelajaran Awal	0.01	Kosinusoidal decay
Resolusi Input	640x640	Standar YOLOv8
Optimizer	SGD	Momentum 0.937
Weight Decay	0.0005	Regulasi anti overfitting
Ambang Kepercayaan	0.5	Keseimbangan presisi-recall
IoU Threshold (NMS)	0.45	Standard NMS post-processing

2.7 Pipeline Deteksi Real-Time

Sistem deteksi real-time dikembangkan menggunakan Python dengan library Ultralytics YOLOv8 untuk inferensi dan OpenCV untuk pemrosesan video. Pipeline beroperasi dalam siklus frame-per-frame meliputi akuisisi frame dari sumber video, pra-pemrosesan frame meliputi pengubahan ukuran ke 640x640 piksel dan normalisasi nilai piksel ke rentang 0 hingga 1, forward pass melalui model YOLOv8s untuk menghasilkan tensor prediksi mentah, post-processing mencakup penerapan ambang batas kepercayaan 0.5 dan Non-Maximum Suppression dengan IoU threshold 0.45 untuk eliminasi deteksi redundan, serta visualisasi hasil deteksi pada frame keluaran dengan bounding box berwarna yang dikodekan per kelas [20].

Logika klasifikasi keselamatan diimplementasikan sebagai lapisan post-processing di atas hasil deteksi model. Jika label kelas helm terdeteksi pada atau di atas wilayah kepala objek orang yang berdekatan, pekerja tersebut diklasifikasikan sebagai SAFETY dan bounding box orang ditandai dengan warna hijau. Sebaliknya, jika label

tidak_helm terdeteksi, pekerja diklasifikasikan sebagai NO SAFETY dan ditandai dengan warna merah. Mekanisme asosiasi bounding box menggunakan kriteria spatial overlap dengan IoU lebih besar dari 0.3.

2.8 Prosedur Evaluasi Kinerja

Evaluasi kinerja model dilaksanakan pada subset pengujian yang tidak pernah diekspos ke model selama fase pelatihan maupun validasi, memastikan estimasi kinerja yang tidak bias. Untuk setiap kategori aktivitas yang dievaluasi, satu frame per 30 frame berurutan diekstraksi dari rekaman video berdurasi satu menit yang diambil pada kecepatan rata-rata 19 fps, menghasilkan sekitar 50 frame sampel per skenario evaluasi. Ground truth untuk setiap frame sampel ditentukan melalui inspeksi visual manual oleh dua penilai independen untuk memastikan reliabilitas interrater [21].

Prediksi model kemudian dibandingkan dengan ground truth yang telah ditetapkan untuk mengisi elemen-elemen confusion matrix, yang menjadi dasar perhitungan metrik evaluasi. Akurasi mengukur proporsi prediksi yang benar dari total prediksi. Presisi mengukur kemurnian prediksi positif model. Recall mengukur kelengkapan deteksi kelas positif. F1-Score merupakan harmonik rata-rata dari presisi dan recall yang memberikan gambaran kinerja seimbang. Dalam konteks pemantauan keselamatan konstruksi, F1-Score menjadi metrik yang paling representatif untuk menilai kinerja keseluruhan sistem karena baik false positive maupun false negative memiliki konsekuensi operasional yang signifikan.

3. Hasil dan Diskusi

3.1 Karakteristik Dataset dan Proses Pelatihan

Dataset akhir yang digunakan dalam penelitian ini terdiri dari total 4.170 anotasi bounding box yang terdistribusi di 7 kelas objek. Distribusi dataset antarkelas menunjukkan keseimbangan yang relatif baik, dengan jumlah anotasi per kelas berkisar antara 540 untuk kelas Tidak Memakai Helm hingga 700 untuk kelas Orang, perbedaan yang tidak cukup besar untuk menyebabkan bias kelas yang signifikan. Setelah augmentasi, subset pelatihan yang efektif mengandung lebih dari 11.000 anotasi, yang cukup untuk pelatihan model YOLOv8s menggunakan pendekatan transfer learning [22].

Proses pelatihan selama 200 epoch berjalan selama sekitar 4,5 jam pada GPU NVIDIA T4 di Google Colaboratory. Kurva training loss menunjukkan penurunan yang progresif dan stabil, dari nilai awal sekitar 3.8 pada epoch pertama hingga nilai konvergen sekitar 0.9 pada epoch ke-150 dan seterusnya. Validation loss mengikuti tren serupa dengan sedikit divergensi pada epoch-epoch akhir yang mengindikasikan awal overfitting marginal, namun tidak cukup signifikan untuk memerlukan penghentian dini. Model terbaik (best.pt) dicatat pada epoch ke-178 berdasarkan validation mAP50 tertinggi sebesar 0.784, yang digunakan sebagai model final untuk seluruh evaluasi selanjutnya.

3.2 Hasil Confusion Matrix dan Metrik Kinerja

Secara keseluruhan, hasil evaluasi menunjukkan heterogenitas yang nyata dalam kinerja deteksi di antara kelas-kelas yang dievaluasi, dengan rentang akurasi yang cukup lebar dari 42% untuk kelas Tidak Memakai Helm hingga 92% untuk kelas Memotong Bata dan Plester Bata. Heterogenitas ini mencerminkan perbedaan tingkat diskriminabilitas visual, konsistensi postur, dan kerumitan gerakan yang spesifik terhadap masing-masing aktivitas. Rata-rata F1-Score yang tidak berbobot di seluruh enam kelas adalah 73%, yang dapat diklasifikasikan sebagai kinerja memadai dalam terminologi evaluasi model machine learning untuk aplikasi pemantauan konstruksi [23].

Tabel 3. Metrik Kinerja Model YOLOv8s di Seluruh Kelas Aktivitas

Kelas	Akurasi	Presisi	Recall	F1-Score
Memasang Bata	56%	85%	62%	72%
Memotong Bata	92%	93%	88%	90%
Mengaduk Semen	82%	77%	92%	84%
Plester Bata	92%	86%	54%	67%
Helm	72%	94%	55%	70%
Tidak Helm	42%	90%	40%	55%

Kelas Memotong Bata mencapai kinerja tertinggi dengan akurasi 92%, presisi 93%, recall 88%, dan F1-Score 90%. Kinerja superior ini secara konsisten dapat diatribusikan kepada karakteristik visual yang sangat khas dan mudah dibedakan dari aktivitas pemotongan bata, yaitu keberadaan alat pemotong angle grinder yang spesifik dan berukuran relatif besar, percikan api dan partikel abu yang menjadi tanda visual yang kuat, postur tubuh yang khas dengan kedua tangan memegang alat di depan tubuh pada ketinggian pinggang hingga dada, serta vibrasi alat yang terekam sebagai blur gerakan yang karakteristik. Kombinasi tanda-tanda visual yang kuat, konsisten, dan beragam ini menghasilkan representasi kelas yang sangat diskriminatif dalam ruang fitur model [24].

Kelas Mengaduk Semen menunjukkan profil kinerja yang menarik dengan akurasi 82%, presisi 77%, recall 92%, dan F1-Score 84%. Nilai recall sebesar 92% merupakan yang tertinggi di antara semua kelas, mengindikasikan bahwa model sangat sensitif dalam mengidentifikasi aktivitas mengaduk semen ketika aktivitas tersebut memang berlangsung. Namun, presisi yang lebih rendah sebesar 77% mengindikasikan adanya false positive yang cukup signifikan. Analisis terhadap frame false positive mengungkapkan bahwa gerakan rotasi tubuh bagian atas yang terjadi dalam konteks aktivitas lain memiliki kemiripan visual dengan gerakan mengaduk semen yang mengakibatkan kebingungan klasifikasi.

Kelas Plester Bata mencapai akurasi 92% yang terlihat tinggi, namun disertai dengan nilai recall yang rendah sebesar 54%, menghasilkan F1-Score 67% yang mencerminkan kinerja yang tidak seimbang. Tidak adanya True Negative dan tingginya False Negative mengindikasikan bahwa model cenderung melewati deteksi aktivitas plesteran meskipun mampu mengonfirmasi deteksi ketika memang memberikan output positif. Fenomena ini kemungkinan disebabkan oleh kemiripan visual antara gerakan sapuan trowel horizontal yang lebar dengan postur berdiri relaksasi pekerja yang beristirahat sejenak [25].

Kelas Memasang Bata menghasilkan akurasi terendah di antara kelas-kelas aktivitas bata sebesar 56%, dengan presisi 85%, recall 62%, dan F1-Score 72%. Kondisi ini kemungkinan mencerminkan distribusi kelas yang tidak seimbang dalam subset evaluasi, di mana model lebih sering diekspos pada contoh positif selama pelatihan dan mengembangkan kecenderungan untuk menebak positif ketika menghadapi frame yang ambigu. Aktivitas memasang bata melibatkan gerakan penempatan bata yang halus dan granular pada jarak dekat dari permukaan dinding, yang dalam representasi 2D citra sering kali menghasilkan postur tubuh yang sulit dibedakan dari postur berdiri nonaktif.

3.3 Analisis Kinerja Deteksi Kepatuhan Helm

Hasil deteksi kepatuhan helm menunjukkan profil kinerja yang asimetris antara kelas helm dan tidak memakai helm. Kelas helm mencapai presisi yang sangat tinggi sebesar 94%, mengindikasikan bahwa ketika model memprediksi helm, prediksi tersebut hampir selalu benar. Namun, recall yang lebih rendah sebesar 55% menunjukkan bahwa model melewati sekitar 45% dari instans helm yang sebenarnya ada. Secara praktis, profil ini berarti sistem akan menghasilkan sangat sedikit alarm palsu keselamatan, namun cukup sering gagal mendeteksi pekerja yang memang menggunakan helm [26].

Sebaliknya, kelas Tidak Memakai Helm menunjukkan performa terburuk secara keseluruhan dengan akurasi 42%, recall 40%, dan F1-Score 55%. Tingginya false negative pada kelas ini merupakan kegagalan kritis dari perspektif keselamatan karena sistem sering gagal mengidentifikasi pekerja yang tidak menggunakan helm, yang justru merupakan kondisi yang paling penting untuk dideteksi. Kegagalan ini secara signifikan membatasi utilitas sistem sebagai alat pemantauan keselamatan yang mandiri. Analisis penyebab mengidentifikasi tiga faktor utama, yaitu heterogenitas visual wilayah kepala manusia tanpa helm yang mencakup beragam warna rambut dan warna kulit, ukuran wilayah kepala yang relatif kecil dalam frame yang diambil dari jarak menengah, dan ketidakseimbangan antara jumlah sampel pelatihan untuk kelas Tidak Memakai Helm dibandingkan dengan kelas Helm [27].

3.4 Analisis Faktor Penyebab Kesalahan Deteksi

Analisis komprehensif terhadap pola kesalahan deteksi mengidentifikasi empat faktor penyebab utama yang secara hierarkis berkontribusi terhadap penurunan kinerja model. Faktor pertama dan paling dominan adalah kemiripan morfologi dan konteks visual antarkelas, yang menciptakan zona ambiguitas dalam ruang representasi fitur model. Aktivitas-aktivitas dengan tanda gerak yang tidak khas atau yang berbagi banyak komponen visual dengan aktivitas lain secara inheren lebih sulit untuk diklasifikasikan dengan akurasi tinggi menggunakan pendekatan

klasifikasi satu tahap berbasis penampilan visual semata. Faktor ini paling nyata memengaruhi kelas memasang bata dan tidak memakai helm [28].

Faktor kedua adalah keterbatasan kualitas dan keragaman dataset pelatihan. Meskipun total 4.170 anotasi merupakan dataset yang cukup memadai untuk fine-tuning model pre-trained, keragaman kondisi perekaman yang terbatas pada satu lokasi dan dua posisi kamera mengakibatkan model tidak terekspos pada cukup banyak variasi kondisi selama pelatihan. Faktor ketiga adalah pengaruh variabilitas kondisi pencahayaan yang tidak dikendalikan selama pengumpulan data. Meskipun variabilitas ini sengaja dipertahankan untuk realisme, inkonsistensi fotometrik yang dihasilkan memperkenalkan variasi dalam representasi visual kelas yang tidak selalu tertangani secara optimal oleh augmentasi data yang diterapkan [29].

Faktor keempat adalah pengaruh jarak kamera dan resolusi efektif objek dalam frame. Pada jarak perekaman yang lebih jauh, objek-objek yang menjadi tanda kelas seperti tangan pekerja yang memasang bata atau kepala pekerja tanpa helm menempati area piksel yang sangat kecil, menghasilkan representasi fitur yang miskin dan kurang diskriminatif. Penelitian dalam domain computer vision konstruksi secara konsisten menunjukkan bahwa peningkatan volume dan keragaman dataset pelatihan merupakan strategi paling efektif untuk meningkatkan generalisasi model [30].

3.5 Perbandingan dengan Penelitian Terdahulu

Tabel 4 menyajikan perbandingan ringkas antara hasil penelitian ini dengan beberapa studi representatif dari literatur yang menggunakan pendekatan serupa. Secara umum, kinerja model pada penelitian ini berada dalam rentang yang konsisten dengan penelitian-penelitian terdahulu yang menggunakan YOLOv8 dan algoritma serupa dalam domain pemantauan konstruksi. Penelitian ini memiliki beberapa keunggulan komparatif dibandingkan dengan studi terdahulu, yaitu integrasi simultan deteksi aktivitas multi-kelas dan pemantauan kepatuhan APD dalam satu model terpadu, fokus pada kategori aktivitas pekerjaan bata yang spesifik dan granular, serta konteks Indonesia yang spesifik dengan karakteristik lingkungan konstruksi yang berbeda dari setting internasional yang mendominasi literatur [31].

Peneliti	Tahun	Algoritma	Domain	Akurasi
Fang et al.	2020	CNN	Deteksi helm CCTV	~88%
Lee et al.	2023	CNN dua langkah	Deteksi helm industri	~90%
Arianto	2023	YOLOv5	APD konstruksi Indonesia	~85%
Barlybayev et al.	2024	YOLOv8	PPE detection benchmark	85-92%
Penelitian ini	2025	YOLOv8s	Aktivitas bata + helm	56-92%

Tabel 4. Perbandingan Kinerja dengan Studi Terdahulu

Di sisi lain, keterbatasan utama dibandingkan dengan beberapa studi terdahulu adalah dataset pelatihan yang lebih kecil dan pengujian yang dilakukan di satu lokasi, yang membatasi klaim generalisasi yang dapat dibuat. Studi oleh Fang et al. dan Lee et al. menggunakan dataset yang lebih besar dengan variasi lokasi dan kondisi pencahayaan yang lebih beragam, sehingga menghasilkan akurasi yang lebih tinggi namun dengan kompleksitas pelatihan yang juga lebih besar. Penelitian ini menunjukkan bahwa meskipun dengan dataset yang relatif terbatas, pendekatan YOLOv8s mampu menghasilkan kinerja yang kompetitif untuk aplikasi pemantauan konstruksi praktis [32].

3.6 Diskusi Implikasi untuk Manajemen Konstruksi

Terlepas dari keterbatasan yang telah diidentifikasi, temuan penelitian ini mengindikasikan bahwa sistem computer vision berbasis YOLOv8 memiliki potensi nyata untuk diintegrasikan sebagai komponen dalam ekosistem pemantauan proyek konstruksi yang lebih cerdas. Dalam konteks operasional nyata, sistem dengan kinerja seperti yang dihasilkan dalam penelitian ini, yaitu F1-Score rata-rata 73% dan presisi helm 94%, dapat berfungsi sebagai alat skrining awal yang mengidentifikasi frame-frame mencurigakan untuk ditinjau lebih lanjut oleh pengawas manusia [33].

Model bisnis yang paling realistis untuk mengadopsi teknologi ini dalam industri konstruksi Indonesia adalah model human-in-the-loop, di mana sistem AI bertindak sebagai filter dan pemroses utama data video yang masif, menghasilkan peringatan yang diprioritaskan untuk ditinjau oleh pengawas manusia. Pendekatan ini

memanfaatkan keunggulan AI dalam kecepatan dan konsistensi pemrosesan data skala besar, sambil tetap mengandalkan penilaian manusia untuk situasi yang ambigu atau berimplikasi tinggi. Integrasi dengan sistem manajemen proyek yang sudah ada, seperti platform pelaporan kemajuan berbasis cloud atau dashboard Building Information Modeling, akan semakin meningkatkan nilai tambah sistem ini bagi kontraktor [34].

4. Kesimpulan

Penelitian ini berhasil mengembangkan, mengimplementasikan, dan mengevaluasi secara empiris sistem computer vision berbasis YOLOv8 untuk pemantauan real-time aktivitas pekerja konstruksi di lokasi konstruksi bangunan aktif di Indonesia. Sistem yang dikembangkan mengintegrasikan pendeteksian dan pengklasifikasian empat kategori aktivitas bata serta pemantauan kepatuhan penggunaan helm keselamatan dalam satu model YOLOv8s terpadu yang dilatih menggunakan dataset yang dianotasi secara khusus dari rekaman lapangan. Berdasarkan hasil evaluasi kuantitatif menggunakan metrik confusion matrix, algoritma YOLOv8 terbukti mampu melakukan klasifikasi real-time beberapa kategori aktivitas pekerja bata dengan tingkat akurasi yang bervariasi, namun umumnya memadai untuk aplikasi pemantauan awal. Khususnya untuk aktivitas dengan tanda gerak yang khas seperti memotong bata yang mencapai F1-Score 90% dan mengaduk semen yang mencapai F1-Score 84%. Kinerja deteksi menunjukkan ketergantungan yang kuat terhadap diskriminabilitas visual kelas aktivitas, dengan aktivitas yang dicirikan oleh gerakan halus atau postur yang tumpang tindih dengan kelas lain menghasilkan metrik kinerja yang lebih rendah. Deteksi kepatuhan helm menunjukkan presisi yang sangat tinggi sebesar 94% untuk kelas helm, namun recall yang lebih rendah mengindikasikan keandalan yang tinggi dalam konfirmasi kepatuhan. Keterbatasan dalam deteksi ketidakpatuhan merupakan kasus yang paling kritis dari perspektif keselamatan kerja. Hasil ini mengarah pada kesimpulan bahwa sistem yang diusulkan paling sesuai diterapkan dalam mode human-in-the-loop yang mengintegrasikan kecepatan pemrosesan AI dengan ketelitian penilaian manusia. Untuk penelitian selanjutnya, disarankan untuk memperluas dataset dengan variasi lokasi, kondisi pencahayaan, dan jenis pekerjaan konstruksi, serta mengeksplorasi arsitektur model yang lebih besar atau pendekatan ensemble untuk meningkatkan kinerja pada kelas aktivitas yang saat ini memiliki akurasi rendah.

Referensi

- [1] A. F. Kineber et al., "Benefits of implementing occupational health and safety management systems for the sustainable construction industry: A systematic literature review," *Sustainability*, vol. 15, no. 17, p. 12697, 2023. <https://doi.org/10.3390/su151712697>
- [2] R. Kunodzia, L. Bikitsha, and R. Haldenwang, "Perceived factors affecting the implementation of occupational health and safety management systems in the South African construction industry," *Safety*, vol. 10, no. 1, p. 5, 2024. <https://doi.org/10.3390/safety10010005>
- [3] R. U. Latief et al., "Labor productivity study in construction projects viewed from influence factors," *Civil Engineering Journal*, vol. 9, no. 3, pp. 583-595, 2023. <https://doi.org/10.28991/CEJ-2023-09-03-07>
- [4] M. A. Musarat et al., "Automated monitoring innovations for efficient and safe construction practices," *Results in Engineering*, vol. 22, p. 102057, 2024. <https://doi.org/10.1016/j.rineng.2024.102057>
- [5] A. Jalil Al-Bayati et al., "PPE non-compliance among construction workers: An assessment of contributing factors utilizing fuzzy theory," *Journal of Safety Research*, vol. 85, pp. 242-253, 2023. <https://doi.org/10.1016/j.jsr.2023.02.008>
- [6] A. M. Vukicevic et al., "A systematic review of computer vision-based personal protective equipment compliance in industry practice," *Artificial Intelligence Review*, vol. 57, no. 12, 2024. <https://doi.org/10.1007/s10462-024-10978-x>
- [7] Y. Chen et al., "YOLO-MS: Rethinking multi-scale representation learning for real-time object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 6, pp. 4240-4252, 2025. <https://doi.org/10.1109/TPAMI.2025.3538473>
- [8] M. Safaldin, N. Zaghdien, and M. Mejdoub, "An improved YOLOv8 to detect moving objects," *IEEE Access*, vol. 12, pp. 59782-59806, 2024. <https://doi.org/10.1109/ACCESS.2024.3393835>
- [9] J. Lee and S. Lee, "Construction site safety management: A computer vision and deep learning approach," *Sensors*, vol. 23, no. 2, p. 944, 2023. <https://doi.org/10.3390/s23020944>
- [10] A. Rauzana, A. Zahrah, and W. Dharma, "Critical delay factors for construction projects in Central Aceh District, Indonesia," *F1000Research*, vol. 11, p. 33, 2022. <https://doi.org/10.12688/f1000research.110024.3>
- [11] Y. Nakanishi, T. Kaneta, and S. Nishino, "A review of monitoring construction equipment in support of construction project management," *Frontiers in Built Environment*, vol. 7, p. 632593, 2022. <https://doi.org/10.3389/fbuil.2021.632593>
- [12] E. Naughton et al., "A structured model for continuous improvement methodology deployment and sustainment: A case study," *Heliyon*, vol. 10, no. 21, p. e40034, 2024. <https://doi.org/10.1016/j.heliyon.2024.e40034>
- [13] J. Li et al., "A review of computer vision-based monitoring approaches for construction workers' work-related behaviors," *IEEE Access*, vol. 12, pp. 7134-7155, 2024. <https://doi.org/10.1109/access.2024.3350773>
- [14] L. Ensslin et al., "Constructivist multi-criteria model to support the management of occupational accident risks in civil construction industry," *PLoS ONE*, vol. 17, no. 7, p. e0270529, 2022. <https://doi.org/10.1371/journal.pone.0270529>
- [15] K. Ibrahim, F. Simpeh, and O. Adebawale, "Benefits and challenges of wearable safety devices in the construction sector," *Smart and Sustainable Built Environment*, 2023. <https://doi.org/10.1108/sasbe-12-2022-0266>
- [16] S. F. A. Zaidi et al., "Vision-based construction safety monitoring utilizing temporal analysis to reduce false alarms," *Buildings*, vol. 14, no. 6, p. 1878, 2024. <https://doi.org/10.3390/buildings14061878>
- [17] W. Fang et al., "Computer vision applications in construction safety assurance," *Automation in Construction*, vol. 110, p. 103013, 2020. <https://doi.org/10.1016/j.autcon.2019.103013>

- [18] S. P. Gonzalez-Sabbagh and A. Robles-Kelly, "A survey on underwater computer vision," *ACM Computing Surveys*, vol. 55, no. 13s, p. 1-35, 2023. <https://doi.org/10.1145/3578516>
- [19] A. Barlybayev et al., "Personal protective equipment detection using YOLOv8 architecture on object detection benchmark datasets: A comparative study," *Cogent Engineering*, vol. 11, no. 1, p. 2333209, 2024. <https://doi.org/10.1080/23311916.2024.2333209>
- [20] U. U. Deshpande et al., "Computer-vision based automatic rider helmet violation detection using NVIDIA TAO toolkit and YOLOv8," *Frontiers in Artificial Intelligence*, vol. 8, p. 1582257, 2025. <https://doi.org/10.3389/frai.2025.1582257>
- [21] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427-437, 2009. <https://doi.org/10.1016/j.ipm.2009.03.002>
- [22] B. F. Spencer, V. Hoskere, and Y. Narazaki, "Advances in computer vision-based civil infrastructure inspection and monitoring," *Engineering*, vol. 5, no. 2, pp. 199-222, 2019. <https://doi.org/10.1016/j.eng.2018.11.030>
- [23] L. Liu et al., "Multi-task intelligent monitoring of construction safety based on computer vision," *Buildings*, vol. 14, no. 8, p. 2429, 2024. <https://doi.org/10.3390/buildings14082429>
- [24] M. I. B. Ahmed et al., "Personal protective equipment detection: A deep-learning-based sustainable approach," *Sustainability*, vol. 15, no. 18, p. 13990, 2023. <https://doi.org/10.3390/su151813990>
- [25] B. I. Arianto, "Implementasi metode YOLO pada deteksi pakaian keselamatan yang lengkap di proyek konstruksi," *Ranah Research*, vol. 6, no. 1, pp. 56-63, 2023.
- [26] Y. Kim and Y. Choi, "Smart helmet-based proximity warning system to improve occupational safety using image sensor and artificial intelligence," *Int. J. Environ. Res. Public Health*, vol. 19, no. 23, p. 16312, 2022. <https://doi.org/10.3390/ijerph192316312>
- [27] L. Shen, B. Lang, and Z. Song, "DS-YOLOv8-based object detection method for remote sensing images," *IEEE Access*, vol. 11, pp. 125122-125137, 2023. <https://doi.org/10.1109/ACCESS.2023.3330844>
- [28] V. K. Reja, K. Varghese, and Q. P. Ha, "Computer vision-based construction progress monitoring," *Automation in Construction*, vol. 138, p. 104245, 2022. <https://doi.org/10.1016/j.autcon.2022.104245>
- [29] L.-W. Lung, Y.-R. Wang, and Y.-S. Chen, "Leveraging deep learning and Internet of Things for dynamic construction site risk management," *Buildings*, vol. 15, no. 8, p. 1325, 2025. <https://doi.org/10.3390/buildings15081325>
- [30] A. D. Rafindadi et al., "Caught-in/between accidents in the construction industry: A systematic review," *Safety*, vol. 11, no. 1, p. 12, 2025. <https://doi.org/10.3390/safety11010012>
- [31] J. Seo et al., "Computer vision techniques for construction safety and health monitoring," *Advanced Engineering Informatics*, vol. 29, no. 2, pp. 239-251, 2015. <https://doi.org/10.1016/j.aei.2015.02.001>
- [32] X. Luo et al., "Towards efficient and objective work sampling: Recognizing workers' activities in site surveillance videos with two-stream convolutional networks," *Automation in Construction*, vol. 94, pp. 360-370, 2018. <https://doi.org/10.1016/j.autcon.2018.05.024>
- [33] W. Fang et al., "Computer vision for automated safety compliance checking in construction: A review," *Automation in Construction*, vol. 129, p. 103838, 2021. <https://doi.org/10.1016/j.autcon.2021.103838>
- [34] B. Widodo, H. A. Armanto, and E. Setyati, "Deteksi pemakaian helm proyek dengan metode convolutional neural network," *Journal of Intelligent System and Computation*, vol. 3, no. 1, pp. 45-52, 2021. <https://doi.org/10.52985/insyst.v3i1.157>