

ANALISIS KEAKURATAN COMPUTER VISION DALAM MENDETEKSI PENGGUNAAN MASKER N95 DAN RESPIRATOR MENGGUNAKAN METODE YOLOv11

A. Surya Dwiputra, and Adlian Jefiza

Jurusan Teknik Elektro, Prodi Teknik Mekatronika, Politeknik Negeri Batam, Batam, Indonesia

^{*)}E-mail: suryahmadd@gmail.com

Abstrak

Pengawasan manual terhadap penggunaan masker N95 dan respirator di lingkungan industri dinilai kurang efisien dan membutuhkan banyak sumber daya. Penelitian ini mengembangkan sistem deteksi otomatis memanfaatkan algoritma YOLOv11. Dataset yang dikumpulkan dan dianotasi terdiri atas 2.360 gambar yang mencakup 3.519 instansi (*instances*) dalam lima kelas (N95, Respirator, *Surgical mask*, *Other mask*, dan *No Mask*). Tahap pra-pemrosesan mencakup pengubahan ukuran gambar menjadi 640×640 piksel serta penerapan augmentasi data berupa *flip*, *exposure*, *noise*, rotasi, dan *blur*. Model dilatih selama 100 epoch dan berhasil mencapai *mean Average Precision* (mAP50) optimal sebesar 95.4%, dengan presisi keseluruhan 93.7% dan *recall* 90.1%. Hasil pengujian secara *real-time* menunjukkan akurasi sebesar 100% pada jarak 50–100 cm serta 98% pada berbagai variasi kondisi pencahayaan. Selain itu, sistem dapat mendeteksi beberapa individu sekaligus secara andal dengan performa optimal pada sudut wajah antara 0°–90°. Sistem ini memberikan solusi efektif dalam mendukung penerapan protokol keselamatan kerja.

Kata kunci: YOLOv11, Deteksi Masker, Respirator, Visi Komputer, GUI, Industrial, Keselamatan

Abstract

Manual monitoring of N95 mask and respirator compliance in industrial environments is often inefficient and resource-intensive. This study presents an automated detection system based on the YOLOv11 algorithm. A custom dataset comprising 2,360 images with 3,519 instances across five target classes (N95, Respirator, *Surgical mask*, *Other mask*, and *No Mask*) was collected and annotated. Data preprocessing involved resizing images to 640×640 pixels and applying augmentation techniques, including flipping, exposure adjustment, noise injection, rotation, and blurring. Trained over 100 epochs, the model achieved a mean Average Precision (mAP50) of 95.4%, alongside an overall precision of 93.7% and a recall of 90.1%. *Real-time* evaluation demonstrated 100% accuracy at detection distances of 50–100 cm and 98% accuracy across varying lighting conditions. Furthermore, the system reliably detected multiple individuals simultaneously and performed optimally at facial angles from 0° to 90°. Ultimately, the proposed system offers an effective automated solution to reinforce occupational health and safety protocols.

Keywords: YOLOv11, Mask Detection, Respirator, Computer Vision, GUI, Industrial, Safety

1. Pendahuluan

Dalam lanskap industri yang berkembang pesat, khususnya pada sektor pengecatan dan pelapisan (*painting and coating*), jaminan keselamatan kerja merupakan prioritas utama yang tidak dapat ditawar [1]. Para pekerja di lingkungan ini terus-menerus terpapar partikel berbahaya di udara serta uap kimia yang berisiko buruk bagi kesehatan pernapasan dalam jangka panjang [2][3]. Meskipun penggunaan Alat Pelindung Diri (APD) spesifik, seperti masker N95 dan respirator, telah diwajibkan secara ketat, menjaga kepatuhan yang konsisten di lapangan masih menjadi tantangan yang terus dihadapi [4]. Saat ini, pengawasan keselamatan kerja

masih sangat bergantung pada pemantauan manual oleh petugas *Health, Safety, and Environment* (HSE) [5]. Metode konvensional ini tidak hanya membutuhkan banyak sumber daya (*resource-intensive*), tetapi juga rentan terhadap kekeliruan manusia (*human error*), sehingga kerap gagal memberikan perlindungan *real-time* yang dibutuhkan untuk mencegah kecelakaan kerja maupun risiko kesehatan.

Berbagai penelitian terdahulu telah mengeksplorasi penerapan *computer vision* untuk deteksi masker [6], namun mayoritas kerangka kerja tersebut masih berpusat pada konteks pandemi global dengan objek sasaran terbatas pada masker medis standar. Penelitian oleh Abbasi et al. [7], Khoramdel et al. [8], dan Liu & Ren [9],

misalnya, telah berhasil mengimplementasikan arsitektur YOLO versi awal (seperti YOLOv3 hingga v5) untuk mendeteksi penggunaan masker harian masyarakat [10]. Namun, model-model tersebut tidak dilatih untuk mengekstraksi fitur geometris kompleks dari masker khusus industri, seperti kontur respirator berfilter atau struktur kaku masker N95.

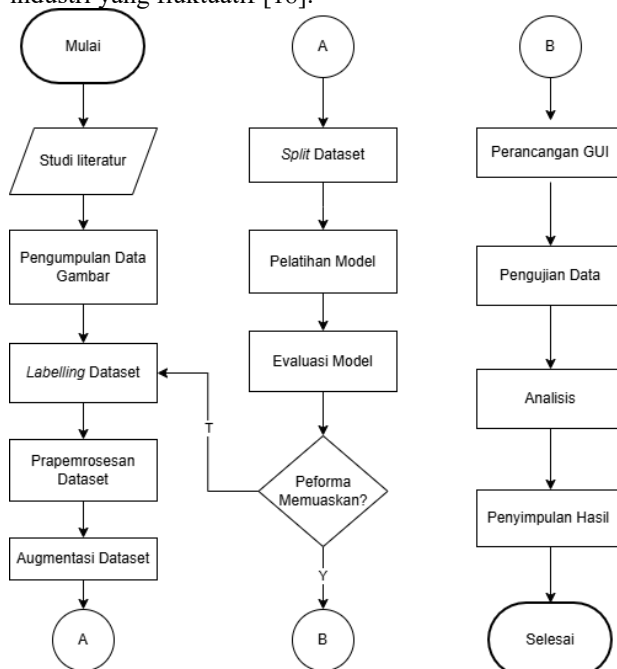
Di sisi lain, upaya peningkatan performa ekstraksi fitur telah dilakukan melalui penerapan algoritma *Convolutional Neural Network* (CNN) ringan oleh Candradewi et al. [11][12] serta YOLOv5 yang dimodifikasi dengan *attention mechanism* oleh Xu et al. [13]. Meskipun kedua pendekatan ini sukses meningkatkan presisi pada kondisi ideal, akurasi model kerap menurun secara signifikan saat dihadapkan pada pencahayaan yang fluktuatif, sudut wajah yang tersembunyi (oklusi), serta kepadatan area kerja [14]. Lebih lanjut, walaupun Toth et al. [6] telah merambah pada deteksi Alat Pelindung Diri (APD) di lingkungan industri baja, penggunaan model ganda (*dual-model*) pada penelitian tersebut menuntut beban komputasi (*computational overhead*) yang terlampau tinggi, sehingga membatasi kelayakan implementasinya untuk pemantauan *real-time* yang efisien.

Menyadari keterbatasan struktural dan operasional dari berbagai penelitian terdahulu, penelitian ini secara eksplisit memposisikan diri untuk menjembatani kesenjangan tersebut melalui pengembangan sistem deteksi yang terspesialisasi untuk lingkungan industri berisiko tinggi. Berbeda dengan penelitian sebelumnya yang rentan terhadap oklusi dan variasi pencahayaan, kerangka kerja ini memanfaatkan iterasi terbaru YOLOv11m. Arsitektur ini dipilih karena memiliki integrasi modul *Cross Stage Partial with Spatial Attention* (C2PSA) yang secara matematis dirancang untuk mempertahankan kesadaran spasial secara optimal, sehingga sangat sensitif terhadap kontur unik N95 dan respirator meski dalam area yang padat [15]. Penelitian ini tidak hanya memperluas domain kelas deteksi pada perlengkapan keselamatan industri berat, tetapi juga menawarkan kebaruan (*novelty*) melalui evaluasi dinamis komprehensif yang menantang model pada variasi jarak (50–400 cm), rentang sudut wajah ekstrem (0°–180°), serta intensitas cahaya redup hingga terang (0–100 lux), guna memastikan sistem siap beroperasi memenuhi standar kepatuhan Health, Safety, and Environment (HSE) secara *real-time* [9][15].

2. Metode

Pengembangan kerangka kerja deteksi keselamatan otomatis ini dilaksanakan melalui alur kerja (*pipeline*) sistematis yang mencakup rekayasa dataset hingga penerapan secara *real-time* [16]. Metodologi ini terbagi ke dalam empat tahapan berurutan: akuisisi dataset *multi-class*, pra-pemrosesan dan augmentasi data, optimasi

arsitektur menggunakan YOLOv11m, serta validasi dinamis melalui *Graphical User Interface* (GUI) kustom berbasis Tkinter [17]. Setiap tahapan dirancang untuk menjamin presisi statistik dan reproduksibilitas operasional di tengah berbagai kendala lingkungan industri yang fluktuatif [18].



Gambar 1. Diagram Alir Perancangan Penelitian

2.1. Akuisisi dan Persiapan Dataset

Dataset yang digunakan dalam penelitian ini diperoleh melalui kombinasi repositori *open-source* dan pengambilan gambar secara langsung yang dilakukan pada berbagai simulasi keselamatan kerja industri, baik di dalam (*indoor*) maupun di luar ruangan (*outdoor*). Skema akuisisi ini menyajikan sudut pandang subjek yang menyeluruh dan tanpa halangan, sehingga pergerakan alami serta beragam variasi posisi penggunaan alat pelindung diri dapat terekam di bawah berbagai kondisi lingkungan yang realistis [19]. Dataset ini secara khusus menargetkan lima kelas objek deteksi: masker N95, respirator, masker medis (*surgical mask*), masker jenis lain (*other mask*), dan kondisi tanpa masker (*no mask*). Untuk komponen pengambilan gambar langsung, data dikumpulkan menggunakan kamera smartphone beresolusi 1280×720 piksel guna merepresentasikan variasi nyata pada intensitas pencahayaan, posisi wajah, serta kompleksitas latar belakang [20]. Melalui proses pengumpulan yang sistematis ini, terkumpul sebanyak 2.360 gambar yang mencakup total 3.519 instansi bounding box awal, dengan rincian: masker N95 (504 instansi), respirator (576 instansi), masker medis (1.184 instansi), masker jenis lain (479 instansi), dan kondisi tanpa masker (776 instansi). Gambar-gambar tersebut selanjutnya dianotasi menggunakan platform Roboflow

dan dibagi secara proporsional ke dalam tiga subset: 70% untuk pelatihan (*training*), 20% untuk validasi (*validation*), dan 10% untuk pengujian (*testing*).



Gambar 2. Contoh Pengambilan Data

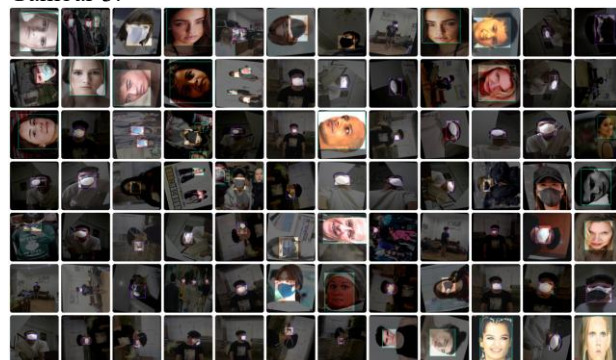
Guna membuktikan stabilitas dan mencegah bias evaluasi akibat pembagian data tunggal, penelitian ini menerapkan skema Stratified 5-Fold Cross-Validation ($K = 5$). Seluruh dataset sebanyak 2.360 gambar dibagi secara acak namun proporsional ke dalam 5 bagian (*fold*). Pada setiap iterasi pengujian, 4 *fold* (80% / 1.888 gambar) digunakan sebagai data pelatihan dan 1 *fold* (20% / 472 gambar) digunakan sebagai data validasi. Proses ini diulang sebanyak 5 kali sehingga seluruh subset data pernah dijadikan data validasi.

2.2. Pra-pemrosesan dan Peningkatan Data

Untuk menstandarisasi input bagi neural network, seluruh gambar melewati tahap pra-pemrosesan (*preprocessing*) dengan mengubah ukurannya secara seragam menjadi 640×640 piksel. Selain itu, teknik augmentasi data diterapkan secara khusus (*exclusively*) hanya pada dataset pelatihan (*training dataset*) [21]. Tahapan ini sangat krusial untuk mencegah terjadinya *overfitting* pada model serta mensimulasikan variabel lingkungan tidak terduga yang sering dijumpai di area industri [22]. Parameter augmentasi yang digunakan meliputi pembalikan horizontal dan vertikal (*horizontal and vertical flips*) untuk mengakomodasi orientasi kamera, penyesuaian pencahayaan (*exposure adjustments*) sebesar $\pm 10\%$ untuk mensimulasikan kondisi pencahayaan yang bervariasi, penambahan *Gaussian noise* (2%) untuk meniru karakteristik sensor kamera berkualitas rendah, rotasi gambar (15°), dan efek buram ringan (*slight blurring*) sebesar 2,5 piksel.

Setelah penerapan teknik augmentasi tersebut, volume dataset meningkat secara signifikan, sehingga memberikan fondasi yang lebih komprehensif untuk pelatihan model yang andal (*robust*). Jumlah total gambar teraugmentasi meningkat menjadi 6.134 berkas, yang mencakup sebanyak 9.146 anotasi *bounding box*. Distribusi instansi akhir dari dataset yang telah diaugmentasi adalah sebagai berikut: masker medis (*Surgical mask*) sebanyak 3.071 instansi (33,6%), tanpa masker (*No Mask*) sebanyak 2.018 instansi (22,1%), respirator (*Respirator*) sebanyak 1.562 instansi (17,1%), masker N95 (*N95 Mask*) sebanyak 1.306 instansi (14,3%),

dan masker jenis lain (*Other mask*) sebanyak 1.189 instansi (13,0%). Contoh tampilan dataset setelah melalui tahap pra-pemrosesan dan augmentasi disajikan pada Gambar 3.



Gambar 3. Visualisasi dataset setelah tahap pra-pemrosesan dan augmentasi data.

2.3. Arsitektur dan Konfigurasi Pelatihan YOLOv11

Sistem yang diusulkan memanfaatkan arsitektur YOLOv11, yang memperkenalkan blok ekstraksi fitur yang ditingkatkan serta *spatial pyramid pooling* yang dioptimalkan dibandingkan dengan generasi terdahulu [23]. Secara struktural, jaringan membagi gambar input menjadi matriks kisi (*grid*) berukuran $S \times S$; jika titik pusat dari objek target berada di dalam sel kisi tertentu, maka sel tersebut bertanggung jawab untuk mengidentifikasi dan menentukan *bounding box* pada objek tersebut [24].

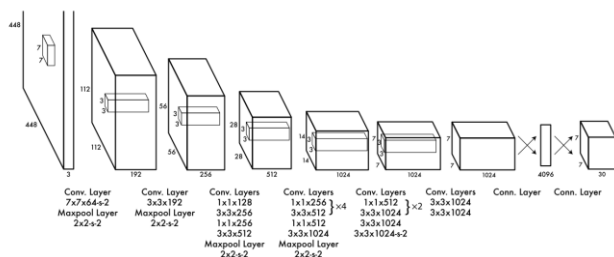
Guna menentukan titik konvergensi paling optimal tempat model mencapai akurasi tinggi tanpa mengalami *overfitting*, tahap pelatihan dilakukan melalui beberapa konfigurasi eksperimen, yaitu 100, 150, 200, dan 500 epoch, dengan bingkai gambar (*input frames*) yang ukurannya diseragamkan menjadi 640×640 piksel dan diproses menggunakan ukuran batch (*batch size*) konstan sebesar 16 gambar [25]. Spesifikasi lengkap perangkat keras (*hardware*) dan perangkat lunak (*software*) yang digunakan dalam penelitian ini dirangkum pada Tabel 1.

Tabel 1. Spesifikasi Lingkungan Perangkat Keras dan Perangkat Lunak

Kategori	Komponen / Perangkat	Spesifikasi Lengkap
Hardware	Central Processing Unit (CPU)	Intel(R) Xeon(R) CPU @ 2.20GHz (Virtual CPU)
	Random Access Memory (RAM)	16 GB System RAM
	Graphics Processing Unit (GPU)	NVIDIA Tesla T4 (VRAM 16 GB GDDR6)

Software	Operating System (OS)	Linux (Google Colab Cloud Environment)
	Programming Language	Python 3.10+
	Core Framework	PyTorch (GPU-accelerated)
	Computer Vision Library	Ultralytics YOLOv11 Ecosystem
	Integrated Development Environment	Visual Studio Code & Google Colab

Proses optimasi model YOLOv11m mengandalkan pengoptimasi *Stochastic Gradient Descent* (SGD) yang dikonfigurasi dengan *momentum* sebesar 0,937 dan *weight decay* sebesar 0,0005 untuk mencegah terjadinya *overfitting* [15]. Skema *learning rate* diatur menggunakan pendekatan *Cosine Annealing decay* dengan nilai awal (lr_0) sebesar 0,01. Pada 3 epoch pertama, diterapkan mekanisme *warm-up* dengan *initial momentum* 0,8 guna menstabilkan gradien sebelum memasuki fase pembobotan utama. Skema pelatihan yang seragam dan terstruktur ini memastikan bahwa setiap perbedaan performa yang teramati selama pengujian murni mencerminkan karakteristik bawaan dari model itu sendiri, bukan disebabkan oleh ketidakkonsistenan dalam konfigurasi perangkat lunak [26].



Gambar 4. Struktur YOLO

Algoritma YOLO menerima input berupa gambar dan memanfaatkan *deep convolutional neural network* untuk mendeteksi objek di dalamnya secara efisien [27]. Arsitektur YOLOv11 terbagi secara sistematis menjadi tiga komponen utama, yang masing-masing memiliki peran spesifik dalam alur kerja (*pipeline*) ekstraksi fitur dan deteksi objek:

Backbone: Komponen ini mengekstraksi fitur-fitur mendasar dari gambar input dengan memanfaatkan blok C3K2, yang merupakan varian struktur *Cross-Stage Partial* (CSP) yang lebih efisien menggunakan kernel berukuran kecil untuk menggantikan blok C2f lama pada YOLOv8. Selain itu, dua lapisan terakhir dari *backbone* mengintegrasikan *Spatial Pyramid Pooling-Fast* (SPPF) untuk ekstraksi fitur multiskala serta *Cross Stage Partial*

with *Spatial Attention* (C2PSA) guna meningkatkan fokus model pada wilayah-wilayah krusial dalam gambar [15]. **Neck:** Komponen ini menggabungkan (*fuses*) fitur-fitur multiskala yang diekstraksi oleh *backbone* menggunakan kombinasi strategis dari blok C3K2, operasi penggabungan (*concatenation*), lapisan konvolusi, dan *upsampling*. Integrasi C2PSA pada bagian *neck* secara efektif membantu model mempertahankan kesadaran spasial (*spatial awareness*), yang sangat krusial untuk mendeteksi objek dengan berbagai ukuran di lingkungan industri yang fluktuatif.

Head: Komponen ini bertanggung jawab menghasilkan luaran (*output*) akhir, meliputi prediksi kelas, skor keberadaan objek (*objectness scores*), dan koordinat *bounding box*, melalui tiga jalur deteksi terpisah yang dioptimalkan untuk objek berukuran kecil, sedang, dan besar. YOLOv11 menerapkan pendekatan *anchor-free* yang memprediksi titik pusat dan dimensi objek secara langsung tanpa bergantung pada *anchor box* yang telah ditentukan sebelumnya, sehingga menjamin presisi secara *real-time*.

Model dilatih dengan resolusi input 640×640 piksel dan ukuran *batch* (*batch size*) sebesar 16 gambar untuk memaksimalkan efisiensi komputasi GPU Tesla T4. Fungsi kerugian (*loss function*) dievaluasi secara multitugas dengan mengombinasikan *Complete IoU* (CIoU) dan *Distribution Focal Loss* (DFL) untuk presisi lokalisasi *bounding box*, serta *Binary Cross-Entropy* (BCE) untuk ketepatan klasifikasi lima kategori masker. Rincian hyperparameter pelatihan secara lengkap disajikan pada Tabel 2.

Tabel 2. Rincian Hyperparameter Pelatihan YOLOv11m

Parameter Pelatihan	Konfigurasi / Nilai	Detail & Keterangan
Model Architecture	YOLOv11m	Model object detection berbasis CNN dari Ultralytics
Input Image Size	640 x 640 piksel	Dimensi input standar untuk ekstraksi fitur efisien
Epochs	100	Jumlah siklus iterasi pelatihan penuh
Batch Size	16	Jumlah sampel gambar per iterasi (optimal untuk GPU Tesla T4)
Optimizer	SGD (Stochastic Gradient Descent)	Dikonfigurasi dengan Momentum = 0,937 dan Weight Decay = 0,0005
Initial Learning Rate (lr_0)	0,01	Batas laju pemelajaran awal
Final Learning Rate (lrf)	0,01	Faktor pengali akhir untuk learning rate decay ($lr_0 \times lrf$)

Learning Rate Schedule	Cosine Annealing	Menurunkan learning rate secara halus mengikuti kurva kosinus
Warm-up Epochs	3,0 epoch	Fase penstabilan awal (Warm-up Momentum = 0,8; Warm-up Initial lr = 0,1)
Loss Function (Bounding Box)	CIoU + DFL	Kombinasi Complete Intersection over Union dan Distribution Focal Loss
Loss Function (Classification)	BCE (Binary Cross-Entropy)	Evaluasi kesalahan klasifikasi antar-kelas masker

2.4. Evaluasi Metrik dan Skenario Pengujian

Peforma model deteksi objek yang telah dilatih dievaluasi secara komprehensif menggunakan kombinasi indikator kuantitatif statis dan skenario validasi dinamis pada kondisi riil (*real-world scenarios*) [28]. Guna mengevaluasi ketepatan dan kelengkapan hasil prediksi jaringan, digunakan metrik standar computer vision, yang meliputi Presisi (*Precision* - P), *Recall* (R), *F1-Score*, serta *mean Average Precision* pada ambang batas (*threshold*) *Intersection over Union* (IoU) sebesar 0,5 (mAP50).

Presisi (*Precision*): Presisi mengukur proporsi deteksi positif yang benar relatif terhadap seluruh hasil prediksi positif. Secara matematis, presisi dinyatakan sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

***Recall*:** *Recall* mengukur kemampuan sistemik model dalam mengidentifikasi kembali instansi positif benar (*true positive*) dari seluruh objek kebenaran acuan (*ground-truth*) yang ada. Secara matematis, *recall* dihitung sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

***F1-Score*:** *F1-Score* merupakan metrik evaluasi model yang mengukur keseimbangan harmonis antara presisi (*precision*) dan *recall*. Secara matematis, *F1-Score* dihitung sebagai berikut:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

***Frame per Second* (FPS):** digunakan untuk mengukur kecepatan sistem dalam memproses gambar secara *real-*

time. Evaluasi FPS menunjukkan beberapa banyak *frame* yang dapat diproses dalam satu detik (1s) oleh sistem. FPS yang tinggi menunjukkan sistem mampu digunakan untuk pemantauan secara langsung tanpa keterlambatan signifikan.

$$FPS = \frac{N}{T} \quad (4)$$

***Mean Average Precision* (mAP):** *Mean Average Precision* (mAP) digunakan untuk menilai performa keseluruhan model dengan menghitung nilai *Average Precision* (AP) rata-rata di seluruh kelas yang ada pada dataset. Dengan mengombinasikan presisi dan *recall* di berbagai ambang batas (*threshold*), ketiga metrik ini saling melengkapi dan bersifat krusial untuk memastikan bahwa sistem deteksi yang dibangun dapat beroperasi secara akurat.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (5)$$

Selain evaluasi statis di laboratorium ini, model yang telah dioptimalkan diimplementasikan ke dalam *Graphical User Interface* (GUI) kustom yang dikembangkan melalui pustaka Tkinter Python untuk mengevaluasi efisiensi inferensi secara *real-time* melalui umpan kamera langsung (*live webcam feed*) [29]. Penerapan operasional ini menguji secara ketat ketahanan struktural serta batas generalisasi sistem terhadap empat variabel industri yang sangat fluktuatif:

Jarak Fisik (*Physical Distance*): Evaluasi persepsi kedalaman spasial kamera dengan jarak fisik antara subjek dan lensa yang diuji secara spesifik pada rentang 50 cm hingga 400 cm.

Sudut Wajah (*Facial Angles*): Analisis toleransi model terhadap variasi geometris yang berkisar dari 0° (tampak depan langsung) hingga 180° (orientasi profil penuh dan menghadap ke belakang).

Kepadatan Kerumunan (*Crowd Density*): Penilaian kemampuan deteksi banyak orang (*multi-person detection*) dengan mengevaluasi bingkai video aktif yang memuat 1 hingga lebih dari 6 individu secara bersamaan dalam satu bingkai (*frame*) yang sama.

Intensitas Pencahayaan Lingkungan (*Environmental Lighting Intensity*): Pengujian stabilitas ekstraksi fitur pada model di tengah tingkat pencahayaan yang fluktuatif, yang dikategorikan secara sistematis ke dalam simulasi cahaya redup (low-light, 0–50 lux) dan kondisi area kerja yang terang (>100 lux).

Rangkaian kerangka evaluasi terpadu ini menyajikan gambaran multidimensi yang komprehensif mengenai tingkat akurasi, keandalan, dan efisiensi performa model YOLOv11 saat ditransisikan dari tahap pelatihan dataset terisolasi ke dalam lingkungan pengawasan keselamatan otomatis secara *real-time*.

3. Hasil dan Analisa

Bagian ini menyajikan hasil evaluasi empiris terhadap arsitektur YOLOv11m menggunakan dataset keselamatan kerja industri yang terspesialisasi. Model ini dilatih dan divalidasi dalam konfigurasi GPU berbasis *cloud* yang dioptimalkan melalui Google Colab, menggunakan resolusi input 640×640 piksel, ukuran batch (*batch size*) 16, serta siklus pelatihan eksperimental hingga 500 epoch untuk menentukan ambang konvergensi yang paling ideal. Penilaian komprehensif ini mencakup berbagai tingkatan pengukuran kinerja, yang menguji ketepatan dan kelengkapan sistem melalui metrik Presisi (*Precision*), *Recall*, F1-Score, dan *mean Average Precision* (mAP50). Selain parameter laboratorium secara statis, hasil penelitian ini juga menguraikan efisiensi penerapan model dalam *Graphical User Interface* (GUI) kustom berbasis Tkinter untuk pemantauan *webcam* secara *real-time*. Evaluasi tersebut mencakup beberapa plot performa, *confusion matrix* terperinci, serta skenario pengujian fisik secara langsung yang meliputi variasi jarak kamera, sudut wajah, kepadatan kerumunan, dan intensitas pencahayaan lingkungan (*lux*). Secara keseluruhan, penyajian data ini memberikan pemahaman mendalam mengenai bagaimana sistem yang diusulkan mampu menyeimbangkan deteksi berpresisi tinggi dengan kemampuan pengawasan otomatis di lingkungan keselamatan industri yang berisiko tinggi.

3.1. Hasil Deteksi

Tahap deteksi menunjukkan bagaimana arsitektur YOLOv11m yang diusulkan secara akurat mengenali dan mengklasifikasikan berbagai jenis penutup wajah untuk keselamatan kerja industri, meliputi masker N95, respirator, masker medis (*surgical mask*), masker jenis lain (*other mask*), dan kondisi tanpa masker (*no mask*). Di bawah konfigurasi 100 epoch yang optimal, model menghasilkan hasil lokalisasi yang sangat stabil dan konsisten, serta berhasil mengidentifikasi peralatan khusus dengan kesalahan lokalisasi (*localization error*) yang minimal.

Gambar 5 menampilkan luaran prediksi (*prediction outputs*) yang dihasilkan oleh model YOLOv11m pada satu bingkai evaluasi tunggal, sehingga memungkinkan validasi langsung terhadap *presisi bounding box geometris* dan skor kepercayaan (*confidence scores*) yang terlokalisasi. Model ini menghasilkan *bounding box* yang sangat andal dengan jumlah positif palsu (*false positives*) yang lebih sedikit untuk APD industri berat, serta mempertahankan skor kepercayaan di atas 90% bahkan untuk respirator yang berlekuk (*contoured respirators*). Selain itu, optimasi *anchor-free* pada YOLOv11 memungkinkan jaringan untuk menentukan titik pusat struktural yang presisi dari masker industri spesifik secara akurat, sehingga secara signifikan meminimalkan kesalahan klasifikasi antara fitur-fitur wajah yang sangat

mirip serta menjamin batas-batas deteksi yang andal (*robust*) di seluruh kategori target.



Gambar 5. Hasil Deteksi Model YOLOv11m pada Berbagai Kondisi Kepadatan

Bukti visual pada Gambar 5 mengonfirmasi keandalan model di bawah kondisi kepadatan kerumunan yang tinggi. Ketika beberapa subjek berada dalam satu bingkai (*frame*) tunggal, sistem secara simultan mengekstraksi fitur dari area wajah yang terpisah. Pada kuadran kiri, individu dengan masker medis (*surgical mask*) maupun tanpa masker dapat terdeteksi (*bounded*) secara tepat, menunjukkan bahwa kerangka kerja YOLOv11 secara efektif mampu mencegah oklusi fitur (*feature occlusion*) pada skenario banyak orang (*multi-person*).

Model tetap mempertahankan skor kepercayaan (*confidence scores*) yang tinggi (berkisar dari 80% hingga lebih dari 87%) saat mendeteksi peralatan khusus seperti masker N95 dan respirator, sebagaimana ditampilkan pada kuadran kanan. Meskipun dihadapkan pada kompleksitas visual berupa siluet yang saling tumpang-tindih (*overlapping*) dan filter respirator yang berlekuk, lokasi *bounding box* tetap sejajar secara presisi dengan pergeseran (*drifting*) yang minimal.

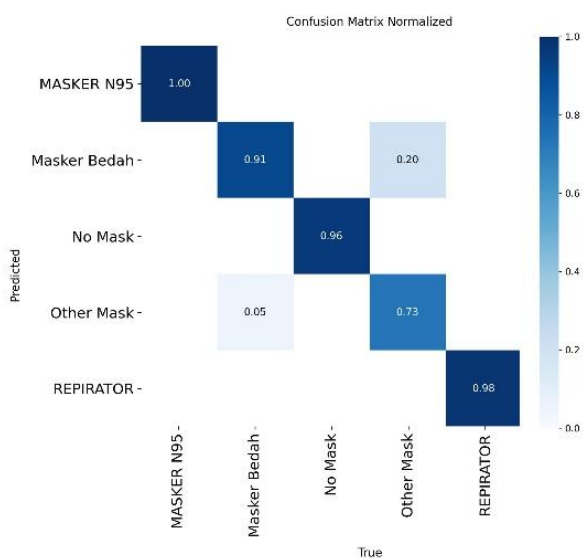
3.2. Pelatihan dan Deteksi Model

Fase pelatihan dan optimasi arsitektur YOLOv11m dievaluasi secara sistematis pada empat konfigurasi *epoch* yang berbeda: 100, 150, 200, dan 500 *epoch*. Pelatihan multi-skenario ini dirancang untuk mengidentifikasi ambang batas matematis yang tepat, di mana jaringan saraf (*neural network*) mencapai bobot fitur yang optimal tanpa mengalami *overfitting*. Evaluasi historis dari proses pelatihan ini didokumentasikan secara kuantitatif pada Tabel 3.

Tabel 3. Perbandingan Kinerja Pelatihan Berdasarkan Epoch

Epoch Configuration	Precision	Recall	mAP50
100 epochs	93.7%	90.1%	95.4%
150 epochs	92%	91.3%	93.6%
200 epochs	93.4%	91.8%	94.1%
500 epochs	93.9%	89.1%	93.2%

Hasil empiris pada Tabel 3 menunjukkan bahwa eksekusi 100 epoch menghasilkan kurva konvergensi terbaik, yang memaksimalkan nilai *mean Average Precision* (mAP50) hingga 95,4%. Ketika batas pelatihan diperpanjang secara signifikan hingga 500 epoch, jaringan menunjukkan tanda-tanda jelas terjadinya *overfitting* ringan, yang dibuktikan oleh penurunan performa mAP50 menjadi 93,2%. Penurunan ini mengonfirmasi bahwa pelatihan iteratif yang terlalu lama menyebabkan model terlalu menghafal noise latar belakang pada data pelatihan, sehingga mengurangi stabilitas generalisasinya pada dataset validasi yang belum pernah dilihat (*unseen validation datasets*). Di bawah konfigurasi bobot 100 epoch yang optimal ini, jaringan *deep learning* menunjukkan sensitivitas yang sangat tinggi dalam mengklasifikasikan Alat Pelindung Diri (APD) industri berat. Nilai *F1-score* untuk kelas N95 Mask dan Respirator masing-masing mencapai angka yang sangat andal, yaitu sebesar 0,993 dan 0,987.

**Gambar 6. Evaluasi Confusion Matrix Model YOLOv11m Yang Dioptimalkan Pada 100 Epoch**

Guna menyelidiki lebih lanjut batas-batas klasifikasi struktural model, verifikasi terperinci terhadap *confusion matrix* telah dilakukan, sebagaimana diilustrasikan pada Gambar 6. Matriks tersebut menunjukkan tingkat *recall* mutlak sebesar 100% untuk masker N95 dan 98% untuk respirator, yang membuktikan bahwa blok ekstraksi fitur (*feature extraction*) yang ditingkatkan berhasil memetakan

kontur geometris unik dari masker-masker khusus tersebut. Sebaliknya, ambiguitas klasifikasi berskala kecil terdeteksi antara kategori *Surgical mask* (masker medis) dan *Other mask* (masker jenis lain) akibat adanya tumpang-tindih (*overlapping*) batas tekstur yang sangat tinggi. Sekitar 20% dari instansi *Other mask* mengalami salah klasifikasi (*misclassified*) sebagai *Surgical mask*, yang membatasi nilai *F1-score standalone* untuk kategori *Other mask* menjadi sebesar 0,814.

Berdasarkan matriks konfusi (*confusion matrix*) pada Gambar 6, terlihat adanya ambiguitas klasifikasi yang cukup signifikan antara kategori *Surgical mask* (masker medis) dan *Other mask* (masker kain/non-medis harian). Terdapat sekitar 20% sampel dari kelas *Other mask* yang salah terprediksi (*misclassified*) sebagai *Surgical mask*. Fenomena ini secara matematis dan visual disebabkan oleh tiga faktor utama:

Pertama adanya kemiripan Geometri dan Fitur Spasial (*High Inter-class Similarity*): Secara spasial, *Surgical mask* dan *Other mask* memiliki kontur lipatan datar (*pleated rectangular structure*) yang menutupi bagian wajah yang identik (area dari tulang hidung hingga dagu). Berbeda dengan *Respirator* atau N95 yang memiliki struktur cembung (*dome-shaped*) kaku yang menonjol, *Surgical mask* dan masker kain umum sering kali memiliki bidang datar serupa saat terpasang pada wajah, sehingga blok konvolusi pada *backbone* mengekstraksi peta fitur (*feature map*) spasial yang sangat mirip.

Kedua terdapat Ketidakseimbangan Jumlah Dataset (*Class Imbalance Bias*): Pada dataset awal, kelas *Surgical mask* mendominasi dengan 1.184 instansi (33,6%), sedangkan kelas *Other mask* hanya memiliki 479 instansi (13,0%). Porsi sampel *Surgical mask* yang jauh lebih besar menyebabkan jaringan *deep learning* mengembangkan bias bobot (*majority class bias*). Ketika model dihadapkan pada gambar dengan fitur tekstur yang kurang jelas (seperti resolusi rendah atau pencahayaan redup), fungsi keputusan (*decision boundary*) model cenderung mengarah pada kelas mayoritas (*Surgical mask*).

Ketiga Hilangnya Fitur Mikro Tekstur pada Augmentasi: Penerapan efek *slight blurring* dan perubahan resolusi input menjadi 640×640 piksel terkadang mengaburkan detail mikro, seperti pola serat bahan *non-woven* sintesis khas masker medis versus serat kain woven/scuba pada *Other mask*.

Guna mengatasi misklasifikasi ini dan meningkatkan *F1-score* kelas *Other mask*, tiga langkah perbaikan teknis diusulkan. Pertama, penerapan *class weighting* melalui skema *Weighted Binary Cross-Entropy* (BCE) atau penyesuaian parameter α pada *Focal Loss* untuk memberikan penalti yang lebih tinggi pada kesalahan prediksi kelas minoritas. Kedua, penyeimbangan proporsi dataset hingga rasio 1:1 antara *Other mask* dan *Surgical mask*, dikombinasikan dengan *targeted data augmentation* seperti *Color Jittering*, *CutMix*, dan *Random Erasing* pada area tali telinga (*earloop*). Ketiga, pengintegrasian modul *attention* beresolusi tinggi seperti *Convolutional Block Attention Module* (CBAM) pada cabang deteksi *Head*

network untuk memperkuat ekstraksi fitur mikro, khususnya pada tekstur lipatan hidung (*nose wire*) dan pembatas tepi masker.

Tabel 4. Hasil Evaluasi Stabilitas Model YOLOv11m Menggunakan 5-Fold Cross-Validation

Fold	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Iteration				
Fold 1	93.5	89.8	95.2	48.5
Fold 2	94.1	90.4	95.6	49.1
Fold 3	93.2	89.5	95.1	48.2
Fold 4	94	90.6	95.7	49.3
Fold 5	93.7	90.1	95.4	48.8
Mean (Rata-rata)	93.7	90.1	95.4	48.8
Std Dev ($\pm$)	0.36	0.45	0.25	0.44

Hasil pengujian Stratified 5-Fold Cross-Validation pada Tabel 4 menunjukkan tingkat stabilitas arsitektur YOLOv11m yang sangat tinggi di seluruh iterasi pembagian data. Model mencatatkan nilai rata-rata mAP@0.5 sebesar 95,4% dengan deviasi standar yang sangat rendah, yaitu $\pm 0,25\%$. Konsistensi ini membuktikan bahwa performa tinggi yang dihasilkan oleh model YOLOv11m murni berasal dari keandalan ekstraksi fitur arsitektur C2PSA dan C3K2, serta bebas dari masalah overfitting maupun kebetulan distribusi data (*data leakage*)

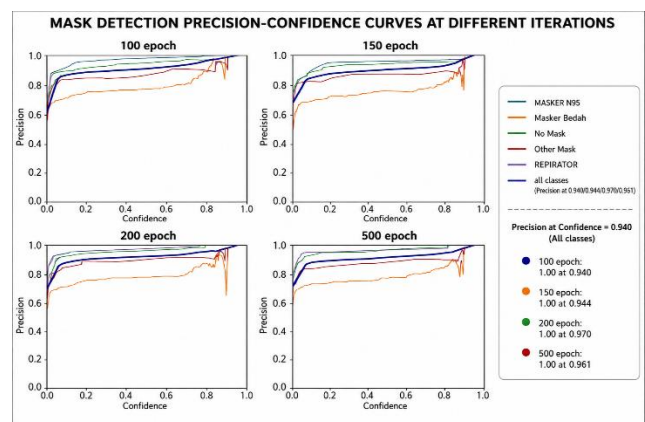
3.3. Confusion Matrix dan Kurva Peforma

Guna menilai lebih lanjut perilaku deteksi spesifik dari model YOLOv11m, serangkaian plot evaluasi dianalisis untuk mengilustrasikan bagaimana jaringan merespons di berbagai ambang batas kepercayaan (*confidence thresholds*) [30]. Visualisasi ini mencakup kurva presisi-kepercayaan (*precision-confidence curve*), kurva recall-kepercayaan (*recall-confidence curve*), kurva F1-kepercayaan (*F1-confidence curve*), serta hubungan presisi-recall (*precision-recall relationship*) secara komprehensif. Secara keseluruhan, kurva-kurva yang dihasilkan ini memberikan pemahaman yang lebih mendalam mengenai bagaimana arsitektur YOLOv11m menyeimbangkan presisi deteksi dan recall selama proses inferensi aktif (*active inference*). Analisis grafis ini menegaskan konsistensi keseluruhan dan keandalan tinggi model saat mengekstraksi fitur untuk kepatuhan keselamatan otomatis sebelum penerapannya ke dalam lingkungan GUI secara *real-time*.

3.3.1. Precision-Confidence Curve

Kurva ini mengilustrasikan bagaimana nilai presisi merespons seiring penyesuaian ambang batas kepercayaan (*confidence threshold*) di berbagai siklus pelatihan. Sebagaimana disajikan pada Gambar 7, jaringan dievaluasi pada 100, 150, 200, dan 500 epoch, di mana seluruh konfigurasi berhasil mencapai puncak presisi sempurna sebesar 1,00 untuk gabungan seluruh kelas. Pada konfigurasi 100 epoch, kerangka kerja ini mencapai skor presisi gabungan sebesar 1,00 pada ambang batas kepercayaan 0,940.

Kurva presisi tersebut menunjukkan bahwa peralatan keselamatan tingkat industri, khususnya kelas MASKER N95 dan RESPIRATOR, secara konsisten mempertahankan posisi tertinggi pada grafik, yang membuktikan rendahnya tingkat deteksi palsu bahkan pada pengaturan kepercayaan yang lebih rendah. Sebaliknya, kelas Masker Bedah membentuk kurva batas bawah di seluruh iterasi sebelum akhirnya meningkat mencapai presisi maksimum pada ambang batas yang lebih tinggi. Secara keseluruhan, tren kenaikan yang stabil di seluruh iterasi membuktikan bahwa sistem yang diusulkan mampu menekan positif palsu (*false positives*) secara efektif, sehingga menjamin penyaringan yang andal untuk kepatuhan keselamatan kerja berbasis otomatisasi.

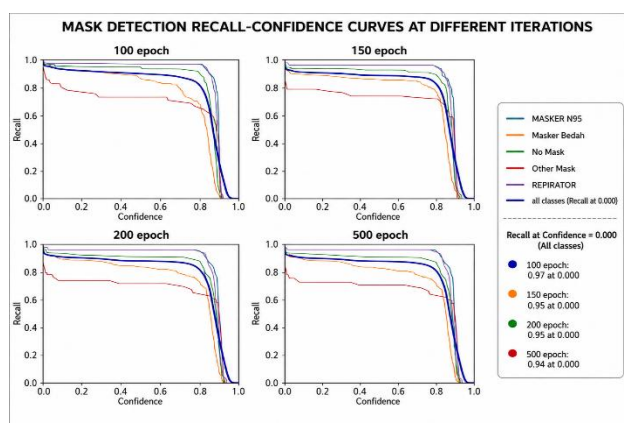


Gambar 7. Perbandingan Kurva Precision-Confidence di Empat Tahap Pelatihan Model yang Berbeda

3.3.2. Recall-Confidence Curve

Kurva ini mencerminkan bagaimana perilaku *recall* model saat ambang batas kepercayaan (*confidence threshold*) disesuaikan di berbagai siklus pelatihan. Sebagaimana disajikan pada Gambar 8, jaringan YOLOv11m dievaluasi melalui empat iterasi, di mana konfigurasi 100 epoch mencatatkan nilai baseline *recall* tertinggi sebesar 0,97 pada nilai 0,000 untuk gabungan seluruh kelas. Kurva historis pada kurva Recall-Kepercayaan (*Recall-Confidence Curve*) menunjukkan bahwa peralatan industri khusus, khususnya kelas RESPIRATOR dan MASKER N95, mempertahankan dataran (*plateau*) *recall* yang

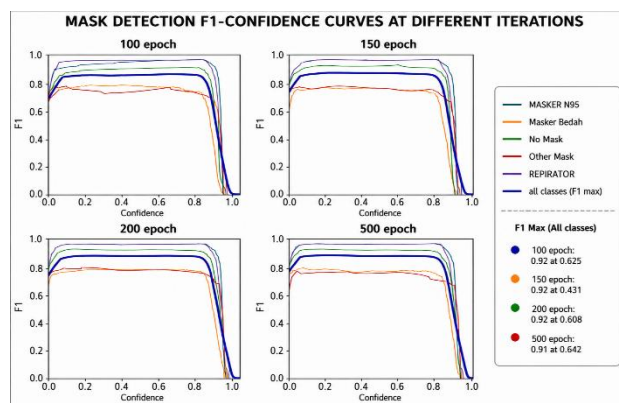
sangat tinggi dan stabil di sepanjang spektrum kepercayaan tingkat rendah hingga menengah, sehingga memastikan APD krusial jarang terlewatkan (*overlooked*). Sebaliknya, kategori *Other mask* membentuk batas performa terendah di seluruh tahapan pelatihan. Tren dataran (*plateau*) yang luas dan stabil sebelum penurunan tajam pada batas kepercayaan ekstrem membuktikan bahwa kerangka kerja yang diusulkan mampu meminimalkan negatif palsu (*false negatives*) secara efektif, sehingga mewujudkan kelengkapan (*completeness*) deteksi yang tinggi untuk pemantauan fasilitas secara otomatis.



Gambar 8. Perbandingan Kurva *Recall-Confidence* di Empat Tahap Pelatihan Model yang Berbeda

3.3.3. *F1-Confidence Curve*

Kurva ini mengilustrasikan bagaimana keseimbangan terpadu antara presisi dan *recall* bergeser seiring perubahan ambang batas kepercayaan (*confidence threshold*) di berbagai siklus pelatihan. Sebagaimana disajikan pada Gambar 9, jaringan dievaluasi melalui empat iterasi, di mana konfigurasi 100 epoch mencapai puncak paling optimal sebesar 0,92 pada ambang batas 0,625 untuk gabungan seluruh kelas. Kurva historis pada Gambar 8 membentuk dataran (*plateau*) yang luas dan tinggi, yang membuktikan bahwa peralatan pelindung industri—khususnya kelas RESPIRATOR dan MASKER N95—mempertahankan stabilitas deteksi yang andal (*robust*) di sepanjang ambang batas kepercayaan operasional yang luas. Sebaliknya, kategori *Other mask* menentukan kurva batas bawah di seluruh tahapan pengujian. Struktur parabola yang memanjang ini mengonfirmasi bahwa sistem mampu mempertahankan keseimbangan klasifikasi dan lokalisasi yang kuat, sekaligus membuktikan kesiapannya untuk implementasi industri secara *real-time*.

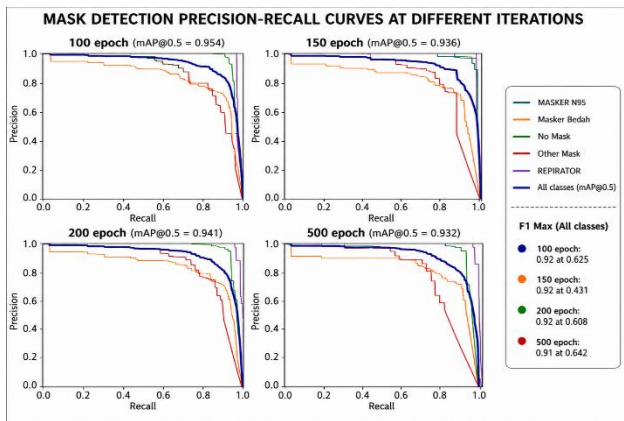


Gambar 9. Perbandingan Kurva *F1-Confidence* di Empat Tahap Pelatihan Model yang Berbeda

3.3.4. *Precision-Recall Curve*

Kurva ini mengilustrasikan hubungan timbal balik (*trade-off*) langsung antara ketepatan dan kelengkapan deteksi di berbagai iterasi. Sebagaimana disajikan pada Gambar 10, performa sistem dipetakan melalui empat tahap pelatihan, di mana kurva paling dominan dicapai pada konfigurasi optimal yang menghasilkan skor keseluruhan sebesar 0,954 mAP@0,5. Grafik tersebut menunjukkan bahwa peralatan pelindung khusus menempati luas area di bawah kurva (*area under the curve—AUC*) tertinggi, dengan kelas MASKER N95 dan RESPIRATOR masing-masing mencatatkan metrik luar biasa sebesar 0,995 dan 0,988.

Dataran (*plateau*) yang tinggi dan stabil pada garis-garis kurva tersebut membuktikan bahwa jaringan secara efektif menekan positif palsu (*false positives*) tanpa mengorbankan integritas *recall*. Sementara itu, kelas standar dan non-industri seperti *No Mask* (0,983), *Other mask* (0,905), dan *Masker Bedah* (0,902) tetap memberikan batas performa yang andal (*robust*). Di seluruh iterasi yang diuji pada Gambar 10, kerangka kerja yang diusulkan secara konsisten mendorong garis biru gabungan ke arah sudut kanan atas, yang membuktikan keandalan strukturalnya untuk pengawasan keselamatan berpresisi tinggi.



Gambar 10. Perbandingan Kurva *Precision-Recall* di Empat Tahap Pelatihan Model yang Berbeda

3.4. Pengujian Sistem Waktu Nyata dan Analisis Lingkungan

Guna memvalidasi kesiapan operasional dari bobot (*weights*) YOLOv11m yang telah dioptimalkan, pengujian lapangan secara dinamis dilakukan melalui *live webcam* stream yang diintegrasikan ke dalam *Graphical User Interface* (GUI) kustom berbasis Tkinter. Ketahanan struktural dan batas-batas generalisasi kerangka kerja ini dievaluasi secara ketat pada empat variabel lingkungan yang sangat fluktuatif: jarak fisik, sudut wajah, kepadatan kerumunan, dan intensitas pencahayaan. Temuan kuantitatif komprehensif dari simulasi *real-time* ini dirangkum dalam Tabel 5.

Tingkat akurasi empiris untuk setiap variabel lingkungan yang diuji dihitung secara sistematis menggunakan formulasi matematis yang membagi total target yang terlokalisasi secara benar dengan total objek yang dijumpai dalam batas bingkai (*frame boundary*) pengujian:

$$Accuracy = \frac{TP}{TI} \times 100\%$$

Dimana:

TP = True Positive (*Successful Detection*)

TI = Total Instance

Sebagaimana dicontohkan pada parameter pemantauan jarak fisik yang dirangkum dalam Tabel 5, di mana seluruh fitur masker target berhasil terdeteksi (*bounded*) di berbagai tingkat kedalaman, penjabaran verifikasi matematisnya adalah sebagai berikut:

$$Accuracy = \frac{21}{21} \times 100\% = 100\%$$

Tabel 5. Kinerja Sistem Waktu Nyata di Bawah Kondisi Lingkungan yang Berubah-ubah

Parameter	Testing Scriptor	Total Instances	Successful Detections	Accuracy Rate (%)
Physical Distance	50 cm – 400 cm	21	21	100

Facial Tracking Angle	0°-180°	24	19	79.1
Multi-Person Density	1 – >10 Persons	102	89	87.2
Lighting Intensity	0 lux – >100 lux	50	49	98

Pertama, persepsi kedalaman spasial diverifikasi pada empat ambang batas (50 cm, 100 cm, 200 cm, dan 400 cm). Sebagaimana dirangkum dalam Tabel 5, sistem mencapai tingkat akurasi sempurna sebesar 100%, dengan berhasil melokalisasi seluruh 21 dari 21 instansi pengujian. Bahkan pada batas maksimum 400 cm, kerangka kerja YOLOv11m yang dioptimalkan tetap mempertahankan *bounding box* yang stabil tanpa pergeseran koordinat (*zero coordinate drifting*) meskipun terjadi penurunan resolusi piksel input.

Kedua, ketahanan terhadap sudut wajah dievaluasi pada spektrum 180° menggunakan 24 instansi pengujian, yang menghasilkan tingkat akurasi sebesar 79,1% (19 deteksi berhasil). Meskipun model beroperasi dengan presisi 100% pada kuadran depan 0° hingga 90°, sebanyak 5 instansi mengalami kegagalan pada sudut di atas 90°. Penurunan performa terlokalisasi ini secara matematis disebabkan oleh oklusi fitur geometris (*geometric feature occlusion*), di mana kontur filter respirator atau klip hidung N95 terhalang sepenuhnya dari garis pandang (*line-of-sight*) lensa.

Ketiga, ekstraksi banyak objek secara simultan (*simultaneous multi-object extraction*) di bawah variasi kepadatan kerumunan (1 hingga 11 orang) mencapai tingkat akurasi 87,2% dengan melokalisasi 89 dari 102 total instansi objek (*total object instances—TI*) secara tepat. Deteksi tetap sempurna (100%) pada konfigurasi jumlah orang yang sedikit (1 hingga 3 orang). Satu kegagalan pelacakan terjadi pada jarak 200 cm selama simulasi 5 orang, di mana kedekatan posisi antarobjek (*high object proximity*) menyebabkan percampuran fitur spasial (*spatial feature blending*) dan tumpang-tindih batas (*overlapping boundaries*).

Keempat, adaptasi pencahayaan disimulasikan pada kondisi cahaya redup (*low-light*, 0–50 lux) dan terang (>100 lux), dengan mencatatkan tingkat akurasi 98,0% (49 dari 50 deteksi berhasil). Hanya 1 negatif palsu (*false negative*) yang tercatat pada ambang batas cahaya redup krisis, yang membuktikan bahwa filter pemajanan (*exposure*) dan keaburan (*slight blurring*) yang diterapkan selama augmentasi data secara efektif melindungi jaringan di 98% skenario lingkungan yang fluktuatif.

Kelima, Pengujian kecepatan deteksi dilakukan untuk memvalidasi kinerja model secara *real-time* menggunakan *input webcam* dengan resolusi 384x640 piksel.

$$FPS = \frac{652}{31,74} = 20,54$$

Pada skenario pengujian ini, sistem berhasil memproses total 652 *frame* dalam waktu 31,74 detik, menghasilkan kecepatan rata-rata 20,54 *Frames Per Second* (FPS) dengan waktu inferensi model yang efisien di kisaran 15 hingga 26 milidetik per *frame*. Pencapaian ini membuktikan bahwa penyesuaian hyperparameter pada model YOLOv11 berhasil menekan beban komputasi secara optimal tanpa mengorbankan akurasi. Secara keseluruhan, kecepatan operasional yang stabil di atas batas standar kelancaran 15 FPS ini mengonfirmasi bahwa sistem deteksi sangat responsif dan sepenuhnya layak diimplementasikan untuk pengawasan pemakaian masker N95 dan respirator secara langsung di lingkungan kerja

4. Kesimpulan

Penelitian ini berhasil menyajikan evaluasi komprehensif terhadap arsitektur YOLOv11m yang diintegrasikan ke dalam *Graphical User Interface* (GUI) kustom berbasis Tkinter untuk pengawasan penutup wajah otomatis di lingkungan industri. Hasil empiris menunjukkan bahwa jaringan mencapai konvergensi optimalnya pada ambang batas 100 *epoch*, dengan meraih nilai *mean Average Precision* (mAP50) yang unggul sebesar 95,4%, presisi keseluruhan sebesar 93,7%, serta sensitivitas yang sangat tinggi terhadap APD khusus, yang dibuktikan oleh nilai *F1-score* sempurna untuk kelas *N95 Mask* (0,993) dan *Respirator* (0,987).

Pengujian lapangan secara *real-time* mengonfirmasi kesiapan operasional kerangka kerja yang andal (*robust*), dengan menghasilkan akurasi mutlak 100% di berbagai tingkat kedalaman spasial (50 cm hingga 400 cm) dan skor adaptasi yang tinggi sebesar 98,0% di bawah kondisi pencahayaan yang fluktuatif (0 lux hingga >100 lux). Terlepas dari kontribusi struktural tersebut, sistem ini masih dibatasi oleh oklusi fitur geometris (*geometric feature occlusion*) yang parah pada sudut wajah profil ekstrem di atas 90° serta sedikit penurunan akurasi pelacakan menjadi 87,2% pada konfigurasi banyak orang (*multi-person*) yang sangat padat akibat batas-batas masker yang saling tumpang-tindih.

Penelitian di masa mendatang harus berfokus pada perluasan dataset untuk menyertakan berbagai sudut pandang profil wajah, melakukan uji tolok ukur (*benchmarking*) kerangka kerja pada perangkat keras *edge-computing*, serta mengintegrasikan teknik optimasi model tingkat lanjut seperti pemangkasan (*pruning*) atau kuantisasi (*quantization*) guna lebih mengurangi beban komputasi (*computational overhead*) dan meningkatkan akurasi lokalisasi banyak objek di lingkungan industri yang padat.

Referensi

Jurnal:

- [1] W. Liu, Q. Meng, Z. Li, and X. Hu, "Applications of Computer Vision in Monitoring the Unsafe Behavior of Construction Workers : Current Status," 2021.
- [2] B. Lin, "Face Mask Detection Based on Improved YOLOv8," vol. 3, pp. 365–375, 2024.
- [3] A. Ying, J. Lv, J. Huang, and T. Wang, "strategy for the motor imagery of force intensity variation," no. June, pp. 1–17, 2025, doi: 10.3389/fnins.2025.1591398.
- [4] A. Regli, A. Sommerfeld, P. Thalayasingam, and B. S. von Ungern-Sternberg, "N95 Masks to Protect Health Care Workers: Is the New Fast Fit-Test Protocol Cutting Corners?," *Chest*, vol. 161, no. 6, pp. 1606–1608, 2022, doi: 10.1016/j.chest.2022.01.048.
- [5] K. L. Andrejko *et al.*, "Effectiveness of Face Mask or Respirator Use in Indoor Public Settings for Prevention of SARS-CoV-2 Infection — California, February–December 2021," *MMWR. Morb. Mortal. Wkly. Rep.*, vol. 71, no. 6, pp. 212–216, 2022, doi: 10.15585/mmwr.mm7106e1.
- [6] K. Toth *et al.*, "A Dual - Model Approach to Industrial Safety : Computer Vision for PPE Compliance and Hazard - Zone Monitoring in Steel Production," *Integr. Mater. Manuf. Innov.*, vol. 15, no. 2, pp. 326–340, 2026, doi: 10.1007/s40192-026-00461-6.
- [7] S. Abbasi, H. Abdi, and A. Ahmadi, "A Face-Mask Detection Approach based on YOLO Applied for a New Collected Dataset," *26th Int. Comput. Conf. Comput. Soc. Iran, CSICC 2021*, pp. 1–6, 2021, doi: 10.1109/CSICC52343.2021.9420599.
- [8] J. Khoramdel, S. Hatami, M. Sadedel, M. Engineering, M. Engineering, and M. Engineering, "Wearing face mask detection using deep learning through COVID- 19 pandemic," pp. 1–13.
- [9] R. Liu and Z. Ren, "Application of Yolo on Mask Detection Task," *2021 IEEE 13th Int. Conf. Comput. Res. Dev. ICCRD 2021*, pp. 130–136, 2021, doi: 10.1109/ICCRD51685.2021.9386366.
- [10] N. Jegham, C. Young, M. Abdelatti, and A. Hendawi, "YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions," arXiv preprint arXiv:2412.18652, pp. 1–18, 2024.
- [11] I. Candradewi, B. A. A. S, A. Harjoko, and A. Dharmawan, "Comparative Analysis of Face Mask Detection using Lightweight CNN and Bag of Visual Word-based Classifier for Real-Time Surveillance," vol. 7, no. 1, pp. 210–225, 2026.
- [12] N. Liao and J. Guan, "Multi - scale Convolutional Feature Fusion Network Based on Attention Mechanism for IoT Traffic Classification," *Int. J. Comput. Intell. Syst.*, 2024, doi: 10.1007/s44196-024-00421-y.
- [13] S. Xu, Z. Guo, Y. Liu, J. Fan, and X. Liu, "An Improved Lightweight YOLOv5 Model Based on Attention Mechanism for Face Mask Detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13531 LNCS, pp. 531–543, 2022, doi: 10.1007/978-3-031-15934-3_44.
- [14] H. Lin and J. Yang, "Ensemble cross-stage partial attention network for image classification," no. July 2021, pp. 102–112, 2022, doi: 10.1049/ipr2.12335.
- [15] L. He, Y. Zhou, L. Liu, and J. Ma, "Research and

- Application of YOLOv11-Based Object Segmentation in Intelligent Recognition at Construction Sites,” *Buildings*, vol. 14, no. 12, 2024, doi: 10.3390/buildings14123777.
- [16] L. Li, “Developments in the Built Environment Automated construction monitoring based on computer vision : A comprehensive review,” *Dev. Built Environ.*, vol. 25, no. December 2025, p. 100832, 2026, doi: 10.1016/j.dibe.2025.100832.
- [17] W. Bagye, “Klasifikasi Kualitas Udara Dengan Metode Support Vector,” *JIRE (Jurnal Inform. Rekayasa Elektron.*, vol. 3, no. 2, pp. 187–199, 2020.
- [18] Y. Han, R. Li, B. Wang, L. Ruan, and Q. Chen, “A pseudo-labeling based weakly supervised segmentation method for few-shot texture images,” vol. 238, no. September 2023, 2024.
- [19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.
- [20] R. Gelar Guntara, “Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7,” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 1, pp. 55–60, 2023, doi: 10.47233/jteksis.v5i1.750.
- [21] R. Khanam and M. Hussain, “YOLOv11: An Overview of the Key Architectural Enhancements,” vol. 2024, pp. 1–9, 2024, [Online]. Available: <http://arxiv.org/abs/2410.17725>
- [22] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J. Big Data*, 2019, doi: 10.1186/s40537-019-0197-0.
- [23] S. Nakayama, T. Nishide, T. Horike, T. Kishimoto, and S. Kira, “Evaluation of the Efficiency of Respiratory Protective Equipment based on the Biological Monitoring of Styrene in Fibreglass Reinforced Plastics Industries,” *J. Occup. Health*, vol. 46, no. 2, pp. 132–140, 2004, doi: 10.1539/joh.46.132.
- [24] Ultralytics, “Ultralytics YOLO Evolution: An Overview of YOLOv8, YOLOv10, and YOLOv11 Object Detectors,” *Ultralytics Documentation*, vol. 26, pp. 1–16, 2025.
- [25] E. H. Alkhamash, “A Comparative Analysis of YOLOv9, YOLOv10, YOLOv11 for Smoke and Fire Detection,” *Fire*, vol. 8, no. 1, 2025, doi: 10.3390/fire8010026.
- [26] A. Riansyah and A. H. Mirza, “Pendeteksi Mobil Berdasarkan Merek dan Tipe Menggunakan Algoritma YOLO,” *Smatika J.*, vol. 13, no. 01, pp. 43–52, 2023, doi: 10.32664/smatika.v13i01.719.
- [27] A. Sharma, V. Kumar, and L. Longchamps, “Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species,” *Smart Agric. Technol.*, vol. 9, no. August, p. 100648, 2024, doi: 10.1016/j.atech.2024.100648.
- [28] M. S. Kiran, “Engineering Science and Technology , an International Journal A comprehensive review on YOLO versions for object detection,” vol. 70, no. August, 2025, doi: 10.1016/j.jestch.2025.102161.
- [29] H. Pathak and R. Kapoor, “Computer Vision for UI Testing : Leveraging Image Recognition and AI to Validate Elements and Layouts Computer Vision for UI Testing : Leveraging Image Recognition and AI to Validate Elements and Layouts,” no. April, 2025.
- [30] M. Imran *et al.*, “Personal Protective Equipment Detection : A Deep-Learning-Based Sustainable Approach,” pp. 1–18, 2023.