

Perbandingan Kinerja *YOLOv8n* dan *Faster R-CNN* untuk Deteksi Otomatis Atribut Seragam Sekolah

Argi Ginanjar¹, Adhi Kusnadi², Risnandar³

^{1,2,3} Program Magister Informatika, Universitas Nusa Putra, Sukabumi, Indonesia
argi.ginanjar_mif24@nusaputra.ac.id

Abstrak

Pemeriksaan kelengkapan atribut seragam siswa secara manual masih banyak dilakukan di lingkungan sekolah, khususnya pada sekolah kejuruan dengan jumlah siswa yang besar. Proses tersebut membutuhkan waktu, rentan terhadap subjektivitas, dan sulit terdokumentasi secara konsisten. Penelitian ini bertujuan membandingkan kinerja *YOLOv8n* dan *Faster R-CNN MobileNetV3 320 FPN* dalam mendeteksi atribut seragam sekolah secara otomatis. Dataset yang digunakan terdiri dari 638 citra seragam sekolah dengan lima kelas objek, yaitu *epolet*, *name_tag*, *sabuk*, *sepatu_hitam*, dan *topi*. Kedua model dilatih menggunakan dataset yang sama, kemudian dievaluasi berdasarkan *Intersection over Union (IoU)*, *mean Average Precision (mAP)*, dan *Frames per Second (FPS)* pada konfigurasi eksperimen yang digunakan. Hasil pengujian menunjukkan bahwa *YOLOv8n* memperoleh *IoU* sebesar 0,8713, *mAP50* sebesar 0,9193, *mAP50-95* sebesar 0,6695, dan kecepatan 83,5921 *FPS*. Sementara itu, *Faster R-CNN* memperoleh *IoU* sebesar 0,6714, *mAP50* sebesar 0,6640, *mAP50-95* sebesar 0,3853, dan kecepatan 9,5801 *FPS*. Hasil tersebut mengindikasikan bahwa, pada konfigurasi eksperimen dan subset evaluasi yang digunakan, *YOLOv8n* memiliki keseimbangan yang lebih baik antara akurasi lokalisasi, akurasi deteksi, dan efisiensi inferensi. Dengan demikian, *YOLOv8n* berpotensi lebih sesuai digunakan sebagai dasar pengembangan sistem pemantauan kelengkapan seragam sekolah secara *semi-real-time*.

Kata kunci: atribut seragam sekolah, *Faster R-CNN*, *FPS*, *object detection*, *YOLOv8n*.

Abstract

Manual inspection of students' uniform completeness is still commonly performed in schools, especially in vocational schools with a large number of students. This process is time-consuming, prone to subjectivity, and difficult to document consistently. This study aims to compare the performance of YOLOv8n and Faster R-CNN MobileNetV3 320 FPN in automatically detecting school uniform attributes. The dataset used in this study consists of 638 school uniform images with five object classes, namely epolet, name_tag, sabuk, sepatu_hitam, and topi. Both models were trained using the same dataset and evaluated based on Intersection over Union (IoU), mean Average Precision (mAP), and Frames per Second (FPS) under the experimental configuration used in this study. The experimental results show that YOLOv8n achieved an IoU of 0.8713, mAP50 of 0.9193, mAP50-95 of 0.6695, and an inference speed of 83.5921 FPS. Meanwhile, Faster R-CNN achieved an IoU of 0.6714, mAP50 of 0.6640, mAP50-95 of 0.3853, and an inference speed of 9.5801 FPS. These results indicate that, under the experimental configuration and evaluation subsets used in this study, YOLOv8n provides a better balance between localization accuracy, detection accuracy, and inference efficiency. Therefore, YOLOv8n has the potential to be more suitable as a basis for developing a semi-real-time school uniform completeness monitoring system.

Keywords: *Faster R-CNN*, *FPS*, *object detection*, *school uniform attributes*, *YOLOv8n*.

I. PENDAHULUAN

Pemeriksaan kelengkapan atribut seragam siswa merupakan salah satu kegiatan rutin dalam

penerapan tata tertib sekolah. Pada praktiknya, pemeriksaan tersebut masih banyak dilakukan secara manual oleh guru atau petugas sekolah sehingga membutuhkan waktu, bergantung pada ketelitian

petugas, dan sulit terdokumentasi secara konsisten. Permasalahan ini semakin terasa pada sekolah kejuruan dengan jumlah siswa yang besar, terutama ketika pemeriksaan dilakukan dalam waktu terbatas. Atribut seperti epolet, papan nama, sabuk, sepatu hitam, dan topi merupakan bagian penting dari kelengkapan seragam yang perlu dipantau karena berkaitan dengan kedisiplinan dan keseragaman penampilan siswa. Oleh karena itu, pemanfaatan teknologi *computer vision* melalui deteksi objek dapat menjadi alternatif untuk membantu proses pemantauan atribut seragam secara lebih objektif, cepat, dan terdokumentasi.

Perkembangan metode *object detection* telah memungkinkan sistem komputer mengenali dan melokalisasi objek secara otomatis dari citra digital. Keluarga *YOLO* menjadi salah satu pendekatan deteksi objek yang banyak dikembangkan untuk kebutuhan deteksi cepat dan *real-time*, termasuk melalui pengembangan arsitektur hingga *YOLOv8n* [1], [2]. Di sisi lain, kajian deteksi objek modern menunjukkan bahwa pemilihan model tidak hanya ditentukan oleh akurasi, tetapi juga oleh efisiensi komputasi, kecepatan inferensi, dan kesesuaian terhadap kebutuhan implementasi [3], [4]. Pada konteks atribut seragam, tantangan deteksi juga berkaitan dengan ukuran objek yang relatif kecil, seperti *name tag*, epolet, dan sabuk, karena objek kecil memiliki informasi visual terbatas serta lebih rentan terhadap gangguan latar belakang [5], [6]. Selain itu, penelitian pada domain pakaian menunjukkan bahwa pendekatan *deep learning* dan deteksi objek dapat dimanfaatkan untuk mengenali kategori, gaya, dan atribut visual pakaian [7], [8].

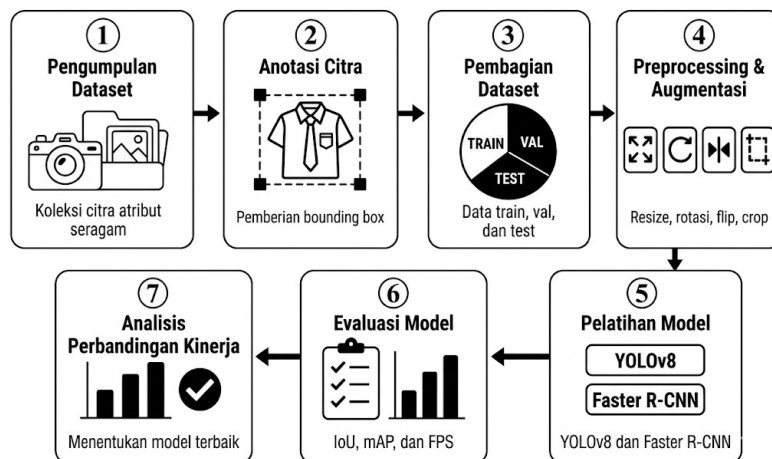
Berdasarkan permasalahan tersebut, penelitian ini bertujuan membandingkan kinerja *YOLOv8n* dan *Faster R-CNN MobileNetV3 320 FPN* dalam mendeteksi atribut seragam sekolah secara otomatis. Dataset yang digunakan terdiri dari lima kelas atribut seragam, yaitu epolet, *name_tag*, sabuk,

sepatu_hitam, dan topi. Kedua model dibandingkan berdasarkan akurasi lokalisasi, akurasi deteksi, dan efisiensi inferensi menggunakan metrik *Intersection over Union (IoU)*, *mAP50*, *mAP50-95*, dan *Frames per Second (FPS)*, yang umum digunakan dalam evaluasi deteksi objek [9], [10]. Kontribusi penelitian ini adalah memberikan analisis komparatif terhadap dua pendekatan deteksi objek pada dataset spesifik atribut seragam sekolah, sehingga dapat menjadi dasar empiris dalam pemilihan model untuk pengembangan sistem pemantauan kelengkapan seragam sekolah secara *semi-real-time*.

II. METODE PENELITIAN

A. Alur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan utama, yaitu pengumpulan dataset, anotasi citra, pembagian dataset, *preprocessing* dan *augmentasi* data, pelatihan model, pengujian model, serta evaluasi kinerja. Model yang dibandingkan dalam penelitian ini adalah *YOLOv8n* sebagai representasi metode *one-stage detector* dan *Faster R-CNN MobileNetV3 320 FPN* sebagai representasi metode *two-stage detector*. Pendekatan *YOLO* dipilih karena keluarga model *YOLO* dirancang untuk melakukan deteksi objek secara langsung dalam satu proses inferensi, sehingga banyak digunakan pada kebutuhan deteksi *real-time* [1], [2]. *Faster R-CNN* digunakan sebagai pembanding karena model ini menerapkan *Region Proposal Network* untuk menghasilkan kandidat wilayah objek sebelum proses klasifikasi dan regresi *bounding box* [11]. Pada implementasi ini, varian *Faster R-CNN MobileNetV3 320 FPN* dipilih karena merupakan model beresolusi rendah dengan *backbone* *MobileNetV3-Large* yang ditujukan untuk kasus penggunaan bergerak atau ringan [11].

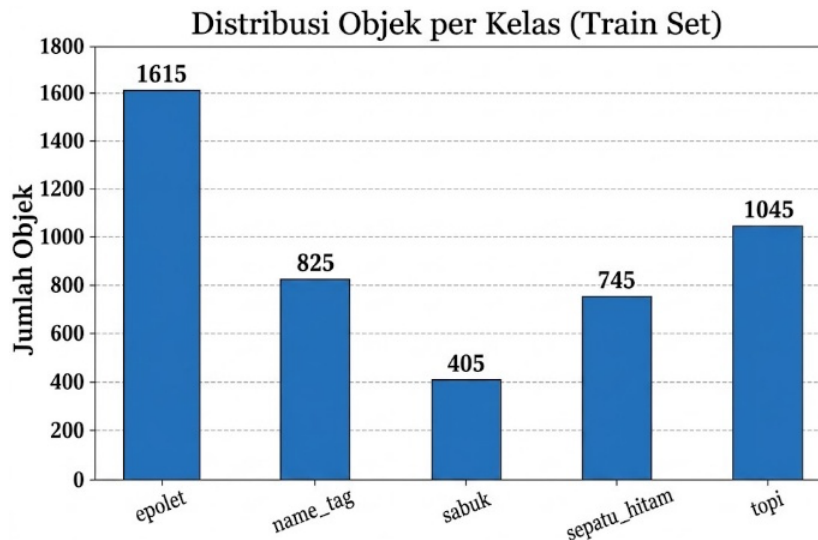


Gambar 1. Alur Penelitian Deteksi Otomatis Atribut Seragam Sekolah

B. Dataset dan Anotasi

Dataset yang digunakan dalam penelitian ini terdiri dari 638 citra seragam sekolah yang diperoleh dari konteks lingkungan sekolah kejuruan. Citra tersebut merepresentasikan atribut seragam siswa dalam berbagai kondisi pengambilan gambar, seperti variasi posisi tubuh, latar belakang, pencahayaan, dan letak atribut pada seragam. Variasi tersebut

diperlukan agar model dapat mempelajari karakteristik objek secara lebih realistis dan tidak hanya bergantung pada kondisi pengambilan gambar yang terlalu terkontrol. Detail jumlah dataset, kelas objek, dan rancangan eksperimen ini berasal dari draf penelitian yang telah disusun sebelumnya. Distribusi jumlah objek pada data latih ditunjukkan pada Gambar 2.



Gambar 2. Distribusi Objek per Kelas pada Data Latih

Objek yang dianotasi terdiri dari lima kelas atribut seragam, yaitu *epolet*, *name_tag*, *sabuk*, *sepatu_hitam*, dan *topi*. Proses anotasi dilakukan menggunakan *bounding box* berbentuk persegi panjang pada setiap atribut yang terlihat dalam citra. Anotasi berbasis *bounding box* digunakan karena kedua model yang dibandingkan merupakan model deteksi objek yang memprediksi lokasi objek dalam bentuk kotak pembatas [12]. Contoh anotasi

bounding box pada masing-masing kelas ditunjukkan pada Gambar 3. Dataset kemudian dibagi menjadi data latih, data validasi, dan data uji dengan rasio 70:20:10. Data latih digunakan untuk proses pembelajaran model, data validasi digunakan untuk memantau performa selama pelatihan, sedangkan data uji digunakan untuk mengukur kinerja akhir model.



Gambar 3. Contoh Anotasi *Bounding Box* pada Setiap Kelas Atribut Seragam

C. Preprocessing dan Augmentasi Data

Tahap *preprocessing* dilakukan untuk menyesuaikan citra dengan kebutuhan masing-masing model. Pada *YOLOv8n*, citra distandarkan ke ukuran 640 × 640 piksel. Sementara itu, *Faster R-CNN MobileNetV3 320 FPN* menggunakan konfigurasi input sesuai *pipeline* model yang digunakan. Selain penyesuaian ukuran citra, dilakukan pula normalisasi piksel agar data masukan lebih stabil selama proses pelatihan.

Augmentasi data diterapkan pada data latih untuk meningkatkan kemampuan generalisasi model terhadap variasi kondisi citra [3], [4]. Teknik *augmentasi* yang digunakan meliputi horizontal flipping, rotasi sudut kecil, penyesuaian kecerahan, dan blur ringan. Augmentasi hanya diterapkan pada data latih, sedangkan data validasi dan data uji tidak diberi *augmentasi* agar hasil evaluasi tetap

mencerminkan performa model pada data yang belum pernah dilihat sebelumnya.

D. Konfigurasi Eksperimen

Penelitian ini menggunakan dua model deteksi objek, yaitu *YOLOv8n* dan *Faster R-CNN MobileNetV3 320 FPN*. *YOLOv8n* dipilih karena memiliki ukuran model yang ringan dan mendukung proses inferensi cepat [2], [12]. *Faster R-CNN MobileNetV3 320 FPN* digunakan sebagai pembanding karena merepresentasikan pendekatan *two-stage detector* yang menggunakan tahapan region proposal sebelum klasifikasi dan regresi *bounding box* [11]. Konfigurasi eksperimen disusun agar kedua model dapat dibandingkan berdasarkan metrik akurasi lokalisasi, akurasi deteksi, dan kecepatan inferensi.

Tabel 1. Konfigurasi Eksperimen

Komponen	<i>YOLOv8n</i>	<i>Faster R-CNN</i>
Arsitektur	<i>YOLOv8n</i>	<i>Faster R-CNN MobileNetV3 320 FPN</i>
Ukuran input	640 × 640 piksel	320 FPN <i>processed input</i>
Jumlah <i>epoch</i>	40	12
Batch size	16 pada GPU / 4 pada CPU	1
Subset evaluasi	Data uji	Data validasi*
Metrik evaluasi	<i>IoU, mAP50, mAP50-95, FPS</i>	<i>IoU, mAP50, mAP50-95, FPS</i>

Faster R-CNN dievaluasi menggunakan data validasi karena struktur label pada data uji tidak dapat dimuat secara identik pada *pipeline* *Faster R-CNN*.

E. Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan menggunakan *Intersection over Union (IoU)*, *mean Average Precision (mAP)*, dan *Frames per Second (FPS)* [9], [10]. *IoU* digunakan untuk mengukur tingkat tumpang tindih antara *bounding box* hasil prediksi dan *bounding box* sebenarnya. Mean Average Precision digunakan untuk mengukur kualitas deteksi berdasarkan hubungan antara precision dan recall. Penggunaan mAP pada beberapa ambang *IoU*, seperti *mAP50* dan *mAP50-95*, umum digunakan dalam evaluasi deteksi objek [9], [10], [13].

IoU dihitung menggunakan Persamaan (1).

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{1}$$

Average Precision dihitung menggunakan Persamaan (2), sedangkan *mean Average Precision* dihitung menggunakan Persamaan (3).

$$AP = \int_0^1 p(r) dr \tag{2}$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \tag{3}$$

Selain akurasi deteksi, penelitian ini juga mengevaluasi kecepatan inferensi menggunakan *FPS*. Metrik ini penting karena sistem yang ditujukan untuk pemantauan seragam sekolah perlu memiliki kemampuan pemrosesan cepat, terutama jika dikembangkan menjadi sistem *semi-real-time*. Kecepatan inferensi dihitung menggunakan Persamaan (4).

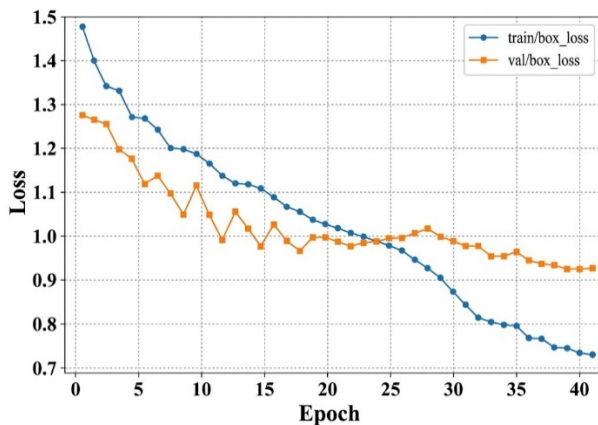
$$FPS = \frac{N_{citra}}{t_{inferensi}} \quad (4)$$

Dengan menggunakan *IoU*, *mAP*, dan *FPS*, evaluasi dalam penelitian ini tidak hanya menilai akurasi deteksi, tetapi juga mempertimbangkan efisiensi model untuk kebutuhan implementasi praktis.

III. HASIL DAN PEMBAHASAN

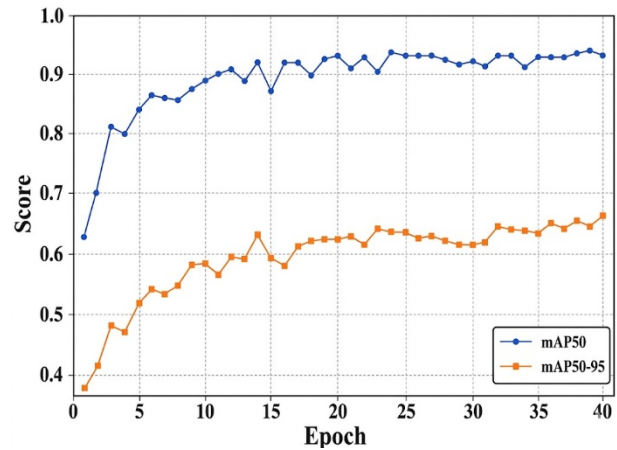
A. Perilaku Pelatihan Model

Proses pelatihan *YOLOv8n* dilakukan selama 40 *epoch* dengan memantau perkembangan nilai *training loss*, *validation loss*, *mAP50*, dan *mAP50-95*. Gambar 4 menunjukkan bahwa nilai *training loss* dan *validation loss* *YOLOv8n* mengalami penurunan secara bertahap selama proses pelatihan. Pola tersebut menunjukkan bahwa model mampu mempelajari karakteristik visual atribut seragam sekolah, seperti *epolet*, *name_tag*, *sabuk*, *sepatu hitam*, dan *topi*. *YOLO* merupakan model deteksi objek yang melakukan prediksi lokasi dan kelas objek secara langsung dalam satu tahapan inferensi [1], [2].



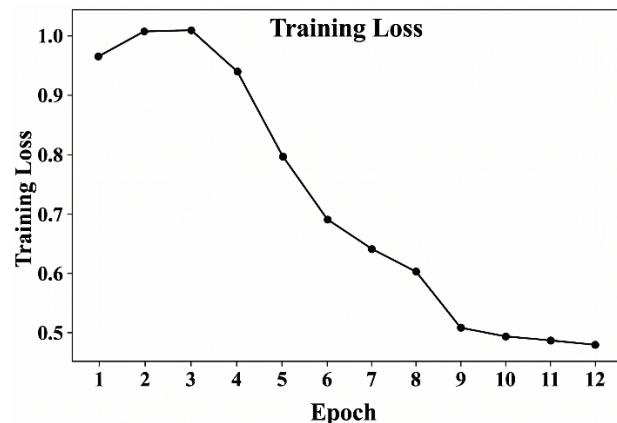
Gambar 4. Perkembangan Nilai *Training Loss* dan *Validation Loss* *YOLOv8n*

Selain penurunan nilai *loss*, perkembangan nilai *mAP* *YOLOv8n* juga menunjukkan pola peningkatan. Gambar 5 memperlihatkan bahwa nilai *mAP50* meningkat cukup cepat pada beberapa *epoch* awal dan kemudian cenderung stabil pada *epoch* berikutnya. Nilai *mAP50-95* juga mengalami peningkatan bertahap, meskipun berada di bawah *mAP50* karena metrik tersebut menggunakan ambang *IoU* yang lebih ketat. Penggunaan *mAP50-95* umum digunakan dalam evaluasi deteksi objek karena menilai performa model pada beberapa ambang *IoU* [9], [10], [13].



Gambar 5. Perkembangan Nilai *mAP50* dan *mAP50-95* *YOLOv8n*

Pada model *Faster R-CNN* *MobileNetV3 320 FPN*, proses pelatihan dilakukan selama 12 *epoch*. Gambar 6 menunjukkan bahwa nilai *training loss* *Faster R-CNN* mengalami penurunan dari *epoch* awal hingga akhir pelatihan. Hal ini menunjukkan bahwa model mampu mempelajari pola objek pada dataset yang digunakan. *Faster R-CNN* menggunakan *Region Proposal Network* untuk menghasilkan kandidat wilayah objek sebelum proses klasifikasi dan regresi *bounding box* dilakukan [11].



Gambar 6. Perkembangan Nilai *Training Loss* *Faster R-CNN*

Berdasarkan grafik pelatihan, kedua model sama-sama menunjukkan proses pembelajaran yang berjalan. Namun, performa akhir kedua model perlu dilihat berdasarkan metrik evaluasi kuantitatif karena penurunan *loss* tidak selalu langsung menunjukkan kemampuan deteksi yang sama pada data evaluasi. Oleh karena itu, analisis berikutnya difokuskan pada perbandingan nilai *IoU*, *mAP50*, *mAP50-95*, dan *FPS*.

B. Hasil Perbandingan Kuantitatif

Hasil perbandingan kinerja *YOLOv8n* dan *Faster R-CNN MobileNetV3 320 FPN* ditunjukkan pada Tabel 2. Metrik yang digunakan meliputi *Intersection over Union (IoU)*, *mAP50*, *mAP50-95*, dan *Frames per Second (FPS)*. *IoU* digunakan untuk mengukur tingkat kesesuaian antara *bounding box* prediksi dan *bounding box ground truth* [9], [13].

Tabel 2. Perbandingan Kinerja YOLOv8n dan Faster R-CNN

Model	Data	IoU	mAP50	mAP50-95	FPS
<i>YOLOv8n</i>	Uji	0,8713	0,9193	0,6695	83,5921
<i>Faster R-CNN MobileNetV3 320 FPN</i>	Val	0,6714	0,6640	0,3853	9,5801

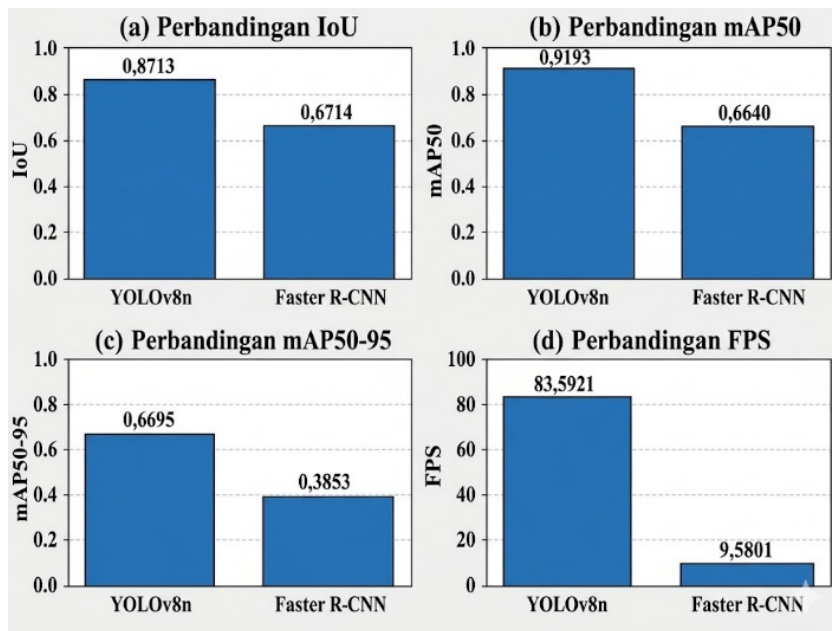
Berdasarkan Tabel 2, *YOLOv8n* memperoleh nilai *IoU* sebesar 0,8713, sedangkan *Faster R-CNN MobileNetV3 320 FPN* memperoleh nilai *IoU* sebesar 0,6714. Hasil tersebut menunjukkan bahwa *YOLOv8n* menghasilkan lokalisasi objek yang lebih baik pada konfigurasi eksperimen yang digunakan. Nilai *IoU* yang lebih tinggi mengindikasikan bahwa posisi *bounding box* hasil prediksi lebih mendekati anotasi sebenarnya.

Pada metrik *mAP50*, *YOLOv8n* memperoleh nilai 0,9193, sedangkan *Faster R-CNN* memperoleh nilai 0,6640. Hasil tersebut menunjukkan bahwa *YOLOv8n* memiliki kemampuan deteksi yang lebih tinggi pada ambang *IoU* 0,50. Pada metrik *mAP50-95*, *YOLOv8n* memperoleh nilai 0,6695, sedangkan *Faster R-CNN* memperoleh nilai 0,3853. Nilai *mAP50-95* yang lebih tinggi menunjukkan bahwa *YOLOv8n* lebih konsisten dalam menghasilkan prediksi pada ambang *IoU* yang lebih ketat.

Perbedaan paling besar terlihat pada metrik *FPS*. *YOLOv8n* memperoleh kecepatan inferensi sebesar 83,5921 *FPS*, sedangkan *Faster R-CNN* memperoleh 9,5801 *FPS*. Dengan demikian, *YOLOv8n* memiliki kecepatan inferensi sekitar 8,73 kali lebih tinggi dibandingkan *Faster R-CNN*. Perbedaan ini menunjukkan bahwa, pada konfigurasi eksperimen ini, *YOLOv8n* lebih potensial untuk skenario pemantauan atribut seragam sekolah yang membutuhkan pemrosesan cepat.

C. Visualisasi Perbandingan Kinerja

Visualisasi perbandingan kinerja *YOLOv8n* dan *Faster R-CNN MobileNetV3 320 FPN* ditunjukkan pada Gambar 7. Grafik tersebut memuat empat metrik utama, yaitu *IoU*, *mAP50*, *mAP50-95*, dan *FPS*.



Gambar 7. Perbandingan Kinerja YOLOv8n dan Faster R-CNN Berdasarkan IoU, mAP50, mAP50-95, dan FPS

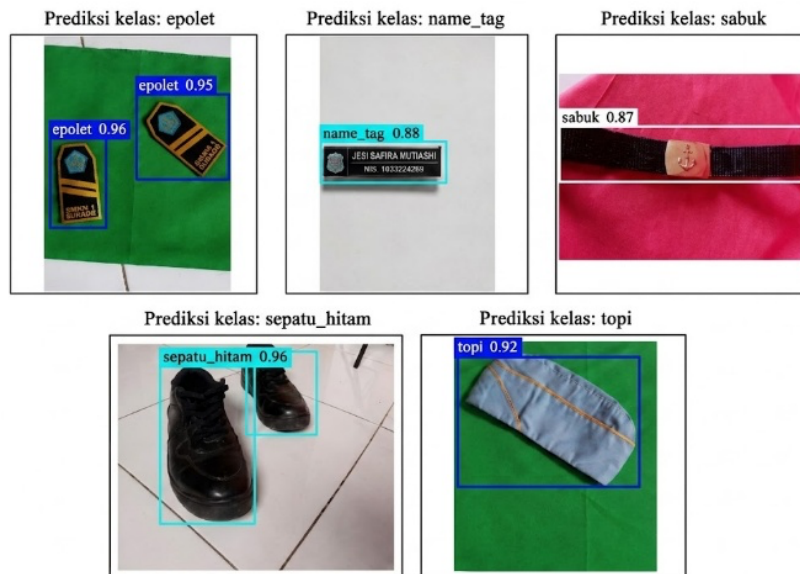
Berdasarkan Gambar 7, *YOLOv8n* menunjukkan kinerja lebih tinggi pada seluruh metrik yang dilaporkan. Namun, hasil ini tetap perlu dibaca dengan mempertimbangkan perbedaan subset evaluasi. *YOLOv8n* dievaluasi pada data uji, sedangkan *Faster R-CNN* dievaluasi pada data

validasi. Oleh karena itu, hasil komparasi ini menunjukkan kecenderungan performa pada konfigurasi eksperimen yang digunakan, bukan klaim mutlak bahwa satu model selalu lebih baik pada semua kondisi.

D. Analisis Hasil Prediksi Visual

Selain evaluasi kuantitatif, penelitian ini juga meninjau hasil prediksi secara visual. Gambar 8 menunjukkan contoh hasil prediksi *YOLOv8n* pada lima kelas atribut seragam sekolah, yaitu epolet,

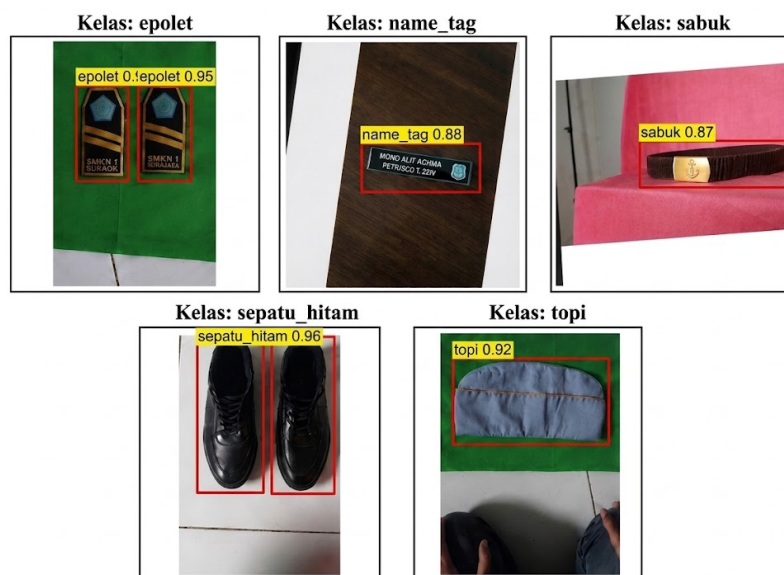
name_tag, sabuk, sepatu hitam, dan topi. Visualisasi tersebut memperlihatkan bahwa *YOLOv8n* mampu menghasilkan *bounding box* pada area objek yang sesuai dengan nilai kepercayaan yang relatif tinggi.



Gambar 8. Contoh Hasil Prediksi *YOLOv8n* pada Setiap Kelas Atribut Seragam

Gambar 9 menunjukkan contoh hasil prediksi *Faster R-CNN* pada kelas atribut yang sama. *Faster R-CNN* juga mampu mendeteksi objek seperti epolet, *name_tag*, sabuk, sepatu hitam, dan topi.

Namun, berdasarkan hasil kuantitatif pada Tabel 2, performa *Faster R-CNN* masih berada di bawah *YOLOv8n* pada seluruh metrik yang dilaporkan.



Gambar 9. Contoh Hasil Prediksi *Faster R-CNN* pada Setiap Kelas Atribut Seragam

Hasil visual pada Gambar 8 dan Gambar 9 menunjukkan bahwa kedua model mampu mengenali objek atribut seragam sekolah. Namun, visualisasi tersebut tidak digunakan sebagai evaluasi kuantitatif per kelas. Visualisasi ini hanya digunakan untuk memberikan gambaran kualitatif mengenai

kemampuan model dalam mendeteksi atribut seragam pada citra.

Deteksi atribut seragam sekolah memiliki tantangan karena beberapa objek berukuran kecil dan memiliki variasi visual yang cukup tinggi. Objek kecil cenderung lebih sulit dideteksi karena

informasi visual yang tersedia lebih terbatas dibandingkan objek berukuran besar [5], [6]. Dalam penelitian ini, tantangan tersebut terlihat pada atribut seperti *name_tag*, epolet, dan sabuk yang hanya menempati sebagian kecil area citra.

Penelitian pada domain pakaian menunjukkan bahwa metode *deep learning* dapat digunakan untuk mengenali atribut visual pada pakaian [7], [8], [14], [15]. Pendekatan berbasis attribute-aware attention juga digunakan dalam klasifikasi fashion untuk meningkatkan representasi atribut visual pada citra pakaian [16]. Penelitian lain juga menunjukkan bahwa pendekatan *YOLO* dapat diterapkan pada pengenalan gaya atau atribut pakaian berbasis citra [8]. Dengan demikian, penelitian ini memperluas penerapan deteksi objek ke konteks yang lebih spesifik, yaitu deteksi otomatis atribut seragam sekolah.

E. Pembahasan

Hasil penelitian menunjukkan bahwa, pada konfigurasi eksperimen yang digunakan, *YOLOv8n* memperoleh kinerja lebih tinggi dibandingkan *Faster R-CNN MobileNetV3 320 FPN* pada seluruh metrik yang dilaporkan. Keunggulan *YOLOv8n* terlihat pada nilai *IoU*, *mAP50*, *mAP50-95*, dan *FPS*. Pada konteks deteksi atribut seragam sekolah, kombinasi antara akurasi deteksi dan kecepatan inferensi menjadi penting karena sistem ditujukan untuk mendukung pemantauan yang bersifat *semi-real-time*.

Perbedaan kinerja antara kedua model dapat dikaitkan dengan karakteristik arsitektur masing-masing [4]. *YOLOv8n* menggunakan pendekatan deteksi satu tahap sehingga proses inferensi dapat dilakukan lebih cepat. *Faster R-CNN* menggunakan pendekatan dua tahap yang melibatkan pembentukan proposal wilayah sebelum klasifikasi objek dilakukan [3], [11]. Pada konfigurasi eksperimen ini, pendekatan *YOLOv8n* lebih menguntungkan karena mampu menghasilkan akurasi yang lebih tinggi sekaligus kecepatan inferensi yang lebih besar.

Hasil penelitian ini juga sejalan dengan kajian *object detection* yang menjelaskan bahwa pemilihan model perlu mempertimbangkan keseimbangan antara akurasi dan efisiensi komputasi [3], [4]. Dalam konteks implementasi, model dengan *FPS* tinggi lebih sesuai untuk sistem yang membutuhkan respons cepat [3], [4], [9]. Oleh karena itu, *YOLOv8n* lebih potensial digunakan sebagai model utama dalam sistem pemantauan atribut seragam sekolah, khususnya pada skenario yang membutuhkan kecepatan inferensi tinggi.

Jika dibandingkan dengan penelitian terkait seragam sekolah, penelitian ini memiliki fokus yang lebih spesifik. Studi sebelumnya menunjukkan

bahwa pendekatan *deep learning* dan *object detection* dapat digunakan untuk identifikasi seragam sekolah maupun deteksi kepatuhan atribut seragam siswa [17], [18]. Penelitian ini tidak hanya mengidentifikasi seragam secara umum, tetapi mendeteksi atribut seragam secara lebih rinci [14], [15], meliputi epolet, *name_tag*, sabuk, sepatu hitam, dan topi. Pendekatan ekstraksi atribut visual pada domain fashion juga menunjukkan pentingnya pengenalan detail atribut pakaian untuk mendukung klasifikasi yang lebih spesifik [19]. Selain itu, penggunaan convolutional neural network pada dataset fashion menunjukkan bahwa pendekatan pembelajaran mendalam dapat digunakan untuk membedakan karakteristik visual pakaian berdasarkan pola citra [20].

Meskipun hasil yang diperoleh menunjukkan keunggulan *YOLOv8n*, penelitian ini memiliki keterbatasan. Pertama, dataset masih berasal dari konteks sekolah tertentu sehingga variasi seragam, pencahayaan, latar belakang, dan pose siswa belum sepenuhnya mewakili kondisi sekolah secara umum. Kedua, subset evaluasi kedua model belum sepenuhnya identik karena *YOLOv8n* dievaluasi pada data uji, sedangkan *Faster R-CNN* dievaluasi pada data validasi. Ketiga, penelitian ini masih berfokus pada deteksi atribut dalam citra, belum mencakup deteksi berbasis video, pelacakan objek, atau integrasi dengan sistem pengambilan keputusan berbasis aturan sekolah.

Berdasarkan keterbatasan tersebut, penelitian lanjutan perlu memperluas dataset dengan melibatkan lebih banyak variasi kondisi pengambilan gambar. Evaluasi berikutnya juga perlu menggunakan subset uji yang sepenuhnya identik untuk semua model agar perbandingan menjadi lebih adil. Selain itu, sistem dapat dikembangkan ke arah deteksi berbasis video dan implementasi pada perangkat edge agar lebih sesuai dengan kebutuhan pemantauan seragam sekolah di lingkungan nyata.

IV. KESIMPULAN

Penelitian ini membandingkan kinerja *YOLOv8n* dan *Faster R-CNN MobileNetV3 320 FPN* untuk deteksi otomatis atribut seragam sekolah pada lima kelas objek, yaitu epolet, *name_tag*, sabuk, sepatu hitam, dan topi. Berdasarkan hasil pengujian, *YOLOv8n* memperoleh *IoU* sebesar 0,8713, *mAP50* sebesar 0,9193, *mAP50-95* sebesar 0,6695, dan kecepatan inferensi 83,5921 *FPS*, sedangkan *Faster R-CNN* memperoleh *IoU* sebesar 0,6714, *mAP50* sebesar 0,6640, *mAP50-95* sebesar 0,3853, dan kecepatan inferensi 9,5801 *FPS*. Hasil tersebut menunjukkan bahwa *YOLOv8n* memiliki keseimbangan lebih baik antara akurasi lokalisasi,

akurasi deteksi, dan efisiensi inferensi pada konfigurasi eksperimen yang digunakan. Dengan demikian, *YOLOv8n* berpotensi digunakan sebagai dasar pengembangan sistem pemantauan kelengkapan atribut seragam sekolah secara *semi-real-time*, dengan tetap mempertimbangkan perlunya evaluasi lanjutan menggunakan subset uji yang sepenuhnya identik. Penelitian selanjutnya disarankan menggunakan subset evaluasi yang sepenuhnya identik untuk seluruh model, memperluas variasi dataset, serta mengembangkan pengujian berbasis video dan perangkat edge.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Program Magister Informatika Universitas Nusa Putra serta pihak sekolah yang telah mendukung proses pengumpulan data dan pelaksanaan penelitian ini.

REFERENSI

- [1] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González, “A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS,” *Mach. Learn. Knowl. Extr.*, vol. 5, no. 4, pp. 1680–1716, 2023, doi: 10.3390/make5040083.
- [2] M. L. Ali and Z. Zhang, “The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in *object detection*,” *Computers*, vol. 13, no. 12, p. 336, 2024.
- [3] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, “Object detection in 20 years: A survey,” *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, 2023.
- [4] M. Trigka and E. Dritsas, “A Comprehensive Survey of Machine Learning Techniques and Models for Object Detection,” *Sensors*, vol. 25, no. 1, 2025, doi: 10.3390/s25010214.
- [5] Q. Feng, X. Xu, and Z. Wang, “Deep learning-based small *object detection*: A survey,” *Math. Biosci. Eng.*, vol. 20, no. 4, pp. 6551–6590, 2023.
- [6] A. Aldubaikhi and S. Patel, “Advancements in small-*object detection* (2023–2025): Approaches, datasets, benchmarks, applications, and practical guidance,” *Appl. Sci.*, vol. 15, no. 22, p. 11882, 2025.
- [7] Y.-H. Chang and Y.-Y. Zhang, “Deep learning for clothing style recognition using YOLOv5,” *Micromachines*, vol. 13, no. 10, p. 1678, 2022.
- [8] X. Li, “Attention-Driven Multi-Scale Clothing Detection Using an Enhanced SCS-YOLO Framework,” *Int. J. Intell. Inf. Technol.*, vol. 21, no. 1, pp. 1–19, 2025.
- [9] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. Da Silva, “A comparative analysis of *object detection* metrics with a companion open-source toolkit,” *Electronics*, vol. 10, no. 3, p. 279, 2021.
- [10] S. Paniago, V. Sharma, and J. M. Cañas, “Open source assessment of *deep learning* visual *object detection*,” *Sensors*, vol. 22, no. 12, p. 4575, 2022.
- [11] PyTorch, “fasterrcnn_mobilenet_v3_large_320_fpn,” May 19, 2026, *PyTorch*. [Online]. Available: https://docs.pytorch.org/vision/master/models/generated/torchvision.models.detection.fasterrcnn_mobilenet_v3_large_320_fpn.html
- [12] N. A. Megantara and E. Utami, “Object Detection using YOLOv8: A Systematic Review,” *Sist. J. Sist. Inf.*, vol. 14, no. 3, pp. 1186–1193, 2025.
- [13] H. Zhang and X. Yang, “A Survey on Algorithms and Performance Metrics for Object Detection,” *Int. Core J. Eng.*, vol. 9, no. 9, pp. 133–145, 2023.
- [14] M. S. Amin, C. Wang, and S. Jabeen, “Fashion sub-categories and attributes prediction model using *deep learning*,” *Vis. Comput.*, vol. 39, no. 9, pp. 3851–3864, 2023.
- [15] V. Parekh and K. Shaik, “Multi-Scale Feature Based Fashion Attribute Extraction Using Multi-Task Learning for e-Commerce Applications.,” in *VISIGRAPP (5: VISAPP)*, 2023, pp. 644–651.
- [16] Y. Wan, C. Yan, B. Zhang, and G. Zou, “Learning image representation via attribute-aware attention networks for fashion classification,” in *International Conference on Multimedia Modeling*, Springer, 2022, pp. 69–81.
- [17] A. F. Saputra and Y. Anang, “Development of Student’s Uniform Compliance Detection System Using Real Time Image Recognition at Politeknik Statistika STIS,” in *Proceedings of The International Conference on Data Science and Official Statistics*, 2023, pp. 58–72.
- [18] M. Dody and Y. Yohannes, “Perangkat Lunak Pendeteksi Jenis Seragam Siswa Jenjang Pendidikan Menengah Menggunakan YOLOv8,” *JATISI*, vol. 12, no. 2, 2025.
- [19] V. Parekh, K. Shaik, S. Biswas, and M. Chelliah, “Fine-grained visual attribute extraction from fashion wear,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 3973–3977.
- [20] O. Nocentini, J. Kim, M. Z. Bashir, and F. Cavallo, “Image classification using multiple convolutional neural networks on the fashion-MNIST dataset,” *Sensors*, vol. 22, no. 23, p. 9544, 2022.

