

Optimasi Arsitektur Deep Learning Berbasis YOLO untuk Deteksi Multi-Kelas Tingkat Kematangan Buah

Achmad Rafly Khatami Zain
Fakultas Informatika, S1 Informatika
Telkom University
Bandung, Indonesia
raflyzainn@student.telkomuniversity.ac.id

Wikky Fawwaz Al Maki
Fakultas Informatika
Telkom University
Bandung, Indonesia
wikkyfawwaz@telkomuniversity.ac.id

Abstrak — Penilaian otomatis tingkat kematangan buah mangga merupakan tantangan dalam pertanian presisi karena perbedaan visual antar tingkat kematangan yang bersifat halus, dipengaruhi oleh variasi pencahayaan dan kondisi lingkungan alami. Pendekatan manual cenderung subjektif, sementara metode visi komputer konvensional masih kesulitan menangani kematangan bertingkat dan ketidakseimbangan data. Penelitian ini mengusulkan pengembangan model deteksi tingkat kematangan buah mangga berbasis YOLOv10s melalui studi ablasi sistematis. Model dikembangkan dengan memvariasikan mekanisme fusi fitur multi-skala, mekanisme *attention*, dan fungsi *loss*, serta dievaluasi pada dataset mangga dengan empat tingkat kematangan menggunakan metrik *precision*, *recall*, *F1-score*, dan *mean Average Precision (mAP)*. Hasil eksperimen menunjukkan bahwa kombinasi BiFPN tiga lapisan, Hybrid Attention Transformer, dan fungsi *loss* EIoU memberikan performa terbaik dengan nilai *mAP@0.5* sebesar 0,96 serta keseimbangan *precision* dan *recall* yang baik. Hasil ini menunjukkan bahwa model yang diusulkan berpotensi diterapkan untuk deteksi tingkat kematangan buah mangga secara multi-tahap pada lingkungan pertanian nyata.

Kata kunci— *attention mechanism*, deteksi objek, *feature fusion*, kematangan buah mangga, YOLOv10s,

I. PENDAHULUAN

Penilaian tingkat kematangan buah secara otomatis merupakan bagian penting dalam pertanian presisi karena mendukung penentuan waktu panen dan kualitas hasil pertanian. Metode inspeksi manual bersifat subjektif, memerlukan waktu lama, dan sulit diterapkan secara konsisten pada skala besar. Selain itu, lingkungan kebun alami menghadirkan tantangan seperti variasi pencahayaan, latar belakang kompleks, serta perbedaan warna dan tekstur buah yang bersifat gradual antar tingkat kematangan, sehingga deteksi kematangan *fine-grained* menjadi permasalahan yang menantang [1], [10], [19].

Penelitian ini berfokus pada buah mangga karena memiliki tahapan kematangan visual yang saling berdekatan, di mana perbedaan antar kelas sering kali hanya ditandai oleh perubahan warna dan tekstur yang sangat halus. Kondisi ini menjadikan deteksi kematangan mangga sebagai studi kasus yang relevan pada lingkungan pertanian nyata [19].

Perkembangan *deep learning* mendorong pemanfaatan algoritma YOLO dalam deteksi objek pertanian karena kemampuannya melakukan lokalisasi dan klasifikasi secara *real-time*. Namun, YOLO konvensional masih memiliki keterbatasan dalam membedakan perbedaan visual yang halus, sensitif terhadap variasi pencahayaan, serta kurang optimal pada kondisi distribusi data yang tidak seimbang [2]–[4].

Berbagai pengembangan telah diusulkan, termasuk fusi fitur multi-skala, mekanisme *attention*, dan optimasi fungsi *loss*. YOLOv10 memperkenalkan *feature alignment* dan *NMS-free decoding* yang meningkatkan performa deteksi pada skenario kompleks [6]–[9]. Namun, analisis sistematis terhadap kontribusi masing-masing komponen tersebut, khususnya pada deteksi kematangan buah mangga, masih terbatas.

Berdasarkan hal tersebut, penelitian ini mengusulkan optimisasi YOLOv10s untuk deteksi tingkat kematangan buah mangga dengan empat kelas (*unripe*, *mid-ripe*, *ripe*, dan *overripe*) melalui integrasi fusi fitur multi-skala, *hybrid attention*, dan fungsi *loss* yang sesuai. Evaluasi dilakukan menggunakan studi ablasi sistematis untuk menganalisis kontribusi setiap komponen arsitektur dengan tetap mempertahankan karakteristik ringan dan *real-time*.

II. KAJIAN TEORI

A. Deteksi Kematangan Buah dalam Pertanian Presisi

Penilaian tingkat kematangan buah merupakan aspek penting dalam pertanian presisi karena berpengaruh langsung terhadap kualitas hasil panen dan waktu panen optimal. Pendekatan inspeksi manual cenderung subjektif, tidak efisien, dan sulit diterapkan secara konsisten pada lingkungan kebun dengan kompleksitas visual tinggi [1]. Meskipun computer vision berbasis *deep learning* memungkinkan deteksi kematangan buah secara otomatis, penerapannya masih menghadapi tantangan berupa variasi pencahayaan, latar belakang kompleks, *occlusion*, serta perbedaan visual antar tingkat kematangan yang sangat halus [1], [10].

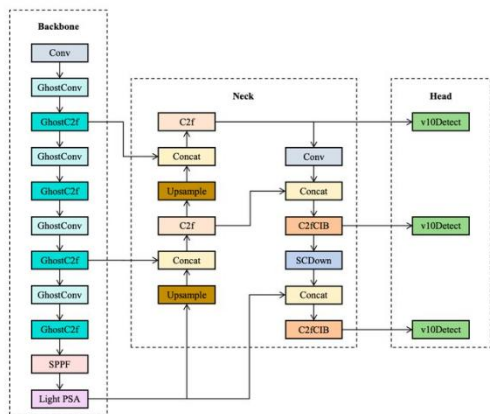
YOLO merupakan arsitektur deteksi objek yang banyak digunakan di bidang pertanian karena kecepatan inferensi yang tinggi dan efisiensi komputasi. Namun, sebagian besar

penelitian sebelumnya masih berfokus pada klasifikasi kematangan berskala kasar atau skenario visual sederhana. YOLOv10 menghadirkan peningkatan melalui stabilitas regresi dan mekanisme NMS-free decoding, tetapi tantangan ambiguitas visual dan ketidakseimbangan kelas pada deteksi kematangan multi-tahap masih memerlukan pengembangan lebih lanjut. Oleh karena itu, YOLOv10 dipilih sebagai fondasi penelitian ini dengan optimasi tambahan untuk mendukung deteksi kematangan buah secara fine-grained.

B. Arsitektur YOLO dan Perkembangannya

You Only Look Once (YOLO) merupakan algoritma deteksi objek berbasis *single-stage detector* yang melakukan lokalisasi dan klasifikasi objek secara simultan dalam satu kali *forward pass*. Dibandingkan metode *two-stage detector* seperti Faster R-CNN, YOLO memiliki keunggulan utama pada kecepatan inferensi dan efisiensi komputasi, sehingga sangat sesuai untuk aplikasi deteksi objek *real-time* di lingkungan pertanian dengan keterbatasan sumber daya [1].

Secara umum, arsitektur YOLO terdiri dari tiga komponen utama, yaitu backbone, neck, dan head. Backbone berfungsi mengekstraksi fitur visual dari citra input, neck menggabungkan fitur multi-skala untuk memperkaya representasi objek, sedangkan head digunakan untuk memprediksi bounding box, skor kepercayaan, dan kelas objek. Struktur modular ini memungkinkan YOLO dikembangkan dan dioptimalkan secara fleksibel sesuai kebutuhan aplikasi. Gambar 1 menunjukkan arsitektur umum YOLO yang terdiri dari tiga komponen utama tersebut.



GAMBAR 1

Diagram Arsitektur Umum YOLO yang terdiri dari Backbone, Neck, dan Head (Diadaptasi dari Liu et al. [22])

Perkembangan arsitektur YOLO terus meningkatkan akurasi dan efisiensi. YOLOv8 memperkenalkan desain yang lebih modular, pendekatan *anchor-free*, serta struktur konvolusi yang lebih ringan, sehingga cocok untuk aplikasi pertanian berbasis perangkat terbatas [2]. Generasi terbaru, YOLOv10, menghadirkan mekanisme *NMS-free decoding* yang menghilangkan proses *Non-Maximum Suppression* pada tahap inferensi. Pendekatan ini meningkatkan stabilitas prediksi dan mengurangi kompleksitas komputasi, khususnya pada objek yang saling berdekatan atau tumpang tindih [7], [8].

Berbagai penelitian menunjukkan bahwa YOLOv10 mampu memberikan performa tinggi pada deteksi objek di lingkungan pertanian yang kompleks, termasuk pada kondisi

occlusion, variasi latar belakang, dan pencahayaan yang tidak merata [7]–[9]. Selain itu, YOLOv10 menyediakan varian ringan seperti YOLOv10s, yang menawarkan keseimbangan antara akurasi dan kecepatan inferensi, sehingga sesuai untuk sistem deteksi kematangan buah secara *real-time*. Berdasarkan pertimbangan tersebut, YOLOv10s dipilih sebagai arsitektur dasar dalam penelitian ini untuk mendukung deteksi tingkat kematangan buah dengan efisiensi komputasi yang tetap terjaga.

C. Mekanisme Attention pada YOLO

Mekanisme attention digunakan untuk meningkatkan fokus model pada fitur visual yang relevan dan menekan pengaruh latar belakang yang kompleks. Pada deteksi objek di lingkungan perkebunan, attention berperan penting dalam menghadapi variasi pencahayaan, *occlusion*, dan perbedaan visual antar objek yang bersifat halus [3].

Pada YOLO, attention umumnya diintegrasikan pada backbone atau neck melalui channel attention dan spatial attention untuk memperkuat representasi fitur. Perkembangan lebih lanjut ditunjukkan oleh Hybrid Attention Transformer (HAT), yang mengombinasikan self-attention untuk menangkap konteks global dengan konvolusi untuk pola lokal. Pendekatan ini efektif dalam membedakan perbedaan visual fine-grained antar tingkat kematangan buah. Oleh karena itu, mekanisme attention digunakan sebagai komponen utama dalam pengembangan YOLOv10s pada penelitian ini.

D. Fungsi Loss pada Deteksi Objek

Fungsi *loss* merupakan komponen penting dalam pelatihan model deteksi objek karena berperan langsung dalam mengarahkan proses optimasi parameter jaringan. Pada YOLO, fungsi *loss* umumnya mencakup kesalahan klasifikasi, regresi *bounding box*, dan prediksi keberadaan objek [10]. Perancangan fungsi *loss* menjadi semakin krusial pada deteksi buah di lingkungan pertanian yang memiliki karakteristik menantang, seperti ukuran objek kecil, *occlusion*, tumpang tindih antar objek, serta ketidakseimbangan distribusi kelas [15], [19].

Pendekatan awal YOLO menggunakan Intersection over Union (IoU) sebagai dasar regresi *bounding box*, namun memiliki keterbatasan ketika prediksi tidak bertumpang tindih dengan *ground truth*. Untuk mengatasi hal tersebut, dikembangkan berbagai varian IoU seperti GIoU, DIoU, dan CIoU yang mempertimbangkan jarak pusat dan rasio aspek objek [21]. Meskipun meningkatkan stabilitas pelatihan, pendekatan ini masih kurang optimal pada objek kecil atau tertutup sebagian.

Beberapa penelitian selanjutnya mengusulkan fungsi *loss* adaptif untuk lingkungan pertanian. Wise-IoU memperkenalkan pembobotan gradien berbasis kualitas sampel untuk meningkatkan stabilitas pelatihan pada kondisi pencahayaan dan latar belakang yang kompleks [4]. Focal-EIoU dikembangkan untuk menangani *class imbalance* dengan menekankan sampel yang sulit diprediksi, sementara Shape-IoU digunakan untuk meningkatkan akurasi regresi pada objek dengan bentuk tidak beraturan [6], [15].

Pada YOLOv10, fungsi *loss* dikembangkan lebih lanjut melalui pendekatan probabilistik dan distribusional, seperti Slide Loss dan Normalized Wasserstein Distance (NWD), yang meningkatkan ketahanan model terhadap noise dan

pergeseran kecil *bounding box* pada objek berukuran kecil [8], [9].

Meskipun berbagai fungsi *loss* telah menunjukkan peningkatan performa, sebagian besar penelitian masih mengevaluasinya secara terpisah tanpa analisis integrasi dengan komponen arsitektur lain, seperti *feature fusion* dan mekanisme *attention*. Oleh karena itu, penelitian ini mengeksplorasi pemilihan fungsi *loss* yang sesuai dengan karakteristik dataset kematangan mangga bertingkat dan mengombinasikannya dengan pengembangan arsitektur untuk meningkatkan stabilitas pelatihan dan akurasi deteksi tanpa mengorbankan efisiensi komputasi.

E. Model YOLO Ringan untuk Lingkungan Pertanian

Implementasi sistem deteksi kematangan buah di lapangan membutuhkan model yang efisien karena sering dijalankan pada perangkat dengan sumber daya terbatas. Meskipun YOLO dikenal cepat, varian berskala besar memiliki kompleksitas tinggi dan kurang sesuai untuk edge computing [5], [20].

YOLOv10s dipilih sebagai model dasar karena menawarkan keseimbangan antara performa dan efisiensi. Dengan dukungan feature alignment dan NMS-free decoding, YOLOv10s lebih stabil dan ringan untuk aplikasi real-time. Model ini kemudian dikembangkan melalui integrasi feature fusion, mekanisme attention, dan fungsi loss yang sesuai untuk meningkatkan akurasi deteksi kematangan mangga pada kondisi lapangan nyata.

III. METODE

A. Gambaran Umum Sistem

Penelitian ini mengusulkan sistem deteksi tingkat kematangan buah mangga berbasis *deep learning* yang mampu melakukan deteksi objek sekaligus klasifikasi tingkat kematangan secara multi-kelas. Sistem dirancang untuk mengklasifikasikan buah mangga ke dalam empat tingkat kematangan, yaitu *unripe*, *mid-ripe*, *ripe*, dan *overripe*, dengan memanfaatkan citra yang diperoleh dari lingkungan kebun alami.

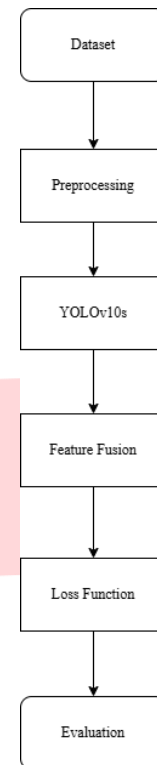
Pengembangan sistem dilakukan secara sistematis, mencakup tahapan pengumpulan dan pemahaman data, pra-pemrosesan dan augmentasi citra, perancangan dan pengembangan arsitektur model, proses pelatihan dan pengujian, serta evaluasi kinerja model. Setiap tahapan dirancang untuk memastikan bahwa sistem mampu menangani perbedaan visual yang bersifat halus antar tingkat kematangan buah, serta tetap robust terhadap variasi kondisi lingkungan seperti perubahan pencahayaan dan latar belakang yang kompleks.

Arsitektur YOLOv10s dipilih sebagai model dasar dalam penelitian ini karena menawarkan keseimbangan yang baik antara akurasi deteksi dan efisiensi komputasi. Karakteristik tersebut menjadikan YOLOv10s sesuai untuk skenario pertanian presisi berbasis *real-time*, khususnya pada implementasi sistem deteksi kematangan buah di lingkungan lapangan dengan keterbatasan sumber daya komputasi.

B. Alur Sistem Deteksi Kematangan Buah

Untuk memberikan gambaran menyeluruh mengenai proses yang diusulkan, Gambar 2 menunjukkan diagram blok sistem deteksi tingkat kematangan buah mangga berbasis

YOLOv10s. Sistem dirancang secara terstruktur, dimulai dari pengolahan dataset hingga tahap evaluasi kinerja model.



GAMBAR 2

Diagram blok sistem deteksi multi-kelas tingkat kematangan buah mangga berbasis YOLOv10s.

Dataset citra buah mangga yang telah dianotasi dalam format YOLO digunakan sebagai data masukan utama. Dataset ini mencakup empat tingkat kematangan buah mangga dan dikumpulkan pada kondisi kebun alami, sehingga memiliki variasi visual yang cukup tinggi. Selanjutnya, citra diproses melalui tahap pra-pemrosesan yang meliputi penyesuaian ukuran, normalisasi nilai piksel, dan augmentasi data untuk meningkatkan keseragaman input serta kemampuan generalisasi model.

Citra hasil pra-pemrosesan kemudian diproses menggunakan arsitektur YOLOv10s sebagai model dasar untuk melakukan ekstraksi fitur dan deteksi objek secara end-to-end. Untuk meningkatkan kualitas representasi fitur, diterapkan mekanisme feature fusion multi-skala yang memungkinkan penggabungan informasi dari berbagai tingkat resolusi, sehingga model lebih efektif dalam mendeteksi buah dengan ukuran bervariasi dan perbedaan kematangan yang bersifat halus.

Pada tahap pelatihan, fungsi loss digunakan untuk mengoptimalkan parameter model dengan mempertimbangkan kesalahan prediksi lokasi dan kelas objek, serta ketidakseimbangan distribusi data antar kelas kematangan. Tahap akhir sistem adalah evaluasi kinerja model menggunakan metrik Precision, Recall, F1-score, dan mean Average Precision (mAP) guna menilai efektivitas sistem dan membandingkan performa antar konfigurasi model yang diuji.

C. Dataset dan Pra-pemrosesan

Dataset yang digunakan pada penelitian ini adalah dataset publik Mango Ripeness Detection yang diperoleh dari Kaggle. Dataset ini terdiri dari 6.106 citra buah mangga yang

telah dianotasi dalam format YOLO dan diklasifikasikan ke dalam empat tingkat kematangan, yaitu *unripe*, *mid-ripe*, *ripe*, dan *overripe*. Seluruh citra dikumpulkan pada kondisi kebun alami, sehingga memiliki variasi visual yang tinggi, seperti perbedaan pencahayaan, latar belakang kompleks, ukuran buah yang beragam, serta adanya *occlusion*. Kondisi ini menjadikan dataset representatif terhadap skenario lapangan nyata.

Dataset dibagi menjadi data pelatihan, validasi, dan pengujian dengan proporsi masing-masing 81,7%, 13,7%, dan 4,6%, sebagaimana ditunjukkan pada Tabel 1. Proporsi data pelatihan dibuat lebih besar untuk mendukung pembelajaran perbedaan visual antar tingkat kematangan yang bersifat halus. Analisis distribusi kelas (Tabel 2) menunjukkan adanya ketidakseimbangan jumlah data antar kelas, khususnya pada kelas *overripe*, yang berpotensi memengaruhi performa model apabila tidak ditangani pada tahap pelatihan.

TABEL 1
DISTRIBUSI DATASET

Pembagian Dataset	Jumlah Gambar	Presentase
Training	4992	81.7
Validasi	839	13.7
Testing	275	4.6
Total	6106	100

TABEL 2
DISTRIBUSI JUMLAH CITRA

Tingkat Kematangan	Jumlah
Unripe	1845
Mid-Ripe	1621
Ripe	1402
Overripe	1238

Contoh citra dari masing-masing tingkat kematangan ditampilkan pada Gambar 3. Perbedaan visual antar kelas terutama ditunjukkan melalui perubahan warna dan tekstur permukaan buah. Namun, transisi visual antara kelas yang berdekatan, seperti *mid-ripe* dan *ripe*, bersifat cukup halus sehingga meningkatkan tingkat kesulitan klasifikasi.



GAMBAR 3
Contoh citra buah mangga

Sebelum digunakan dalam pelatihan, seluruh citra melalui tahap pra-pemrosesan berupa penyesuaian ukuran menjadi 512×512 piksel dan normalisasi nilai piksel ke dalam rentang $[0,1]$. Selain itu, diterapkan augmentasi data untuk meningkatkan keberagaman sampel dan kemampuan generalisasi model, meliputi *horizontal flip*, rotasi kecil ($\pm 15^\circ$), penyesuaian kecerahan dan kontras, perubahan *hue* dan *saturation*, serta *mosaic augmentation*.

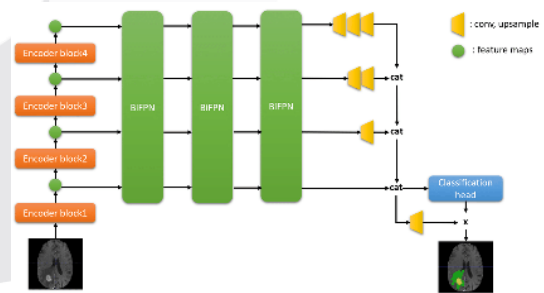
Tahapan pra-pemrosesan dan augmentasi ini bertujuan untuk meningkatkan ketahanan model YOLOv10s terhadap variasi kondisi lingkungan serta perbedaan visual halus antar tingkat kematangan, sehingga performa deteksi menjadi lebih stabil pada penerapan di lapangan nyata.

D. Arsitektur & Model Usulan

Arsitektur YOLOv10s digunakan sebagai model dasar dalam penelitian ini karena menawarkan keseimbangan yang baik antara akurasi dan efisiensi komputasi, sehingga sesuai untuk sistem pertanian presisi berbasis real-time. Meskipun demikian, konfigurasi dasar YOLOv10s masih memiliki keterbatasan dalam membedakan tingkat kematangan buah dengan perbedaan visual yang sangat halus, khususnya pada kelas *mid-ripe* dan *ripe*, serta pada kondisi distribusi data yang tidak seimbang.

Berdasarkan karakteristik dataset, tantangan utama terletak pada kemiripan warna, variasi pencahayaan alami, dan perubahan tekstur yang bersifat gradual, bukan pada *occlusion* yang berat. Oleh karena itu, pengembangan model difokuskan pada peningkatan kualitas representasi fitur dan stabilitas pembelajaran, tanpa menambah kompleksitas arsitektur secara berlebihan. Pendekatan ini juga memungkinkan analisis empiris terhadap kontribusi masing-masing komponen arsitektur secara lebih terkontrol.

Pengembangan utama dilakukan pada mekanisme *feature fusion*. Konfigurasi dasar YOLOv10s menggunakan Path Aggregation Network (PAN), yang kemudian dibandingkan dengan Bidirectional Feature Pyramid Network (BiFPN). BiFPN memungkinkan aliran informasi dua arah dengan bobot fusi adaptif, sehingga penggabungan fitur multi-skala dapat dilakukan secara lebih efektif. Pada penelitian ini, dua konfigurasi BiFPN diuji, yaitu BiFPN dengan dua dan tiga lapisan, untuk menganalisis pengaruh kedalaman fusi fitur terhadap representasi perbedaan visual yang halus. Gambar 4 menunjukkan struktur dasar BiFPN yang digunakan dalam penelitian ini untuk melakukan penggabungan fitur multi-skala secara berulang dengan pembobotan yang efisien.



GAMBAR 4
Arsitektur Bidirectional Feature Pyramid Network (BiFPN) (Diadaptasi dari Nguyen et al. [25])

Setelah *feature fusion*, mekanisme *attention* diintegrasikan untuk memperkuat fitur yang relevan. Tiga konfigurasi diuji, yaitu tanpa *attention*, *Shuffle Attention*, dan Hybrid Attention Transformer (HAT). *Shuffle Attention* digunakan sebagai pendekatan ringan, sedangkan HAT dimanfaatkan untuk menangkap konteks global melalui mekanisme *self-attention*, yang relevan untuk membedakan kelas kematangan dengan karakteristik visual yang sangat mirip.

Modul *attention* ditempatkan setelah tahap *feature fusion* agar bekerja pada representasi fitur multi-skala yang telah diperkaya, sebelum diteruskan ke *detection head*. Seluruh modifikasi dilakukan dengan mempertahankan *backbone* dan *head* YOLOv10s, sehingga perubahan performa dapat dikaitkan langsung dengan mekanisme *feature fusion* dan *attention*.

Berbagai kombinasi konfigurasi model kemudian dievaluasi melalui *ablation study* untuk menganalisis kontribusi masing-masing komponen secara objektif. Model terbaik dipilih berdasarkan keseimbangan antara akurasi deteksi, konsistensi performa antar kelas kematangan, dan efisiensi komputasi, sehingga tetap layak diterapkan pada skenario deteksi kematangan buah mangga di lingkungan lapangan nyata.

E. Fungsi Loss

Pada penelitian ini, fungsi *loss* digunakan sebagai salah satu variabel utama dalam studi ablasi untuk menganalisis pengaruhnya terhadap performa YOLOv10s dalam mendeteksi tingkat kematangan buah mangga. Variasi fungsi *loss* difokuskan pada komponen regresi *bounding box*, mengingat ketelitian lokalisasi memiliki peran penting dalam menentukan keakuratan hasil deteksi secara keseluruhan. Seluruh variasi fungsi *loss* diterapkan pada konfigurasi arsitektur yang sama, sehingga perbedaan performa yang diperoleh dapat dikaitkan secara langsung dengan karakteristik fungsi *loss* yang digunakan.

Fungsi *loss* dasar yang digunakan dalam penelitian ini adalah *Efficient Intersection over Union* (EIoU). EIoU dipilih sebagai *baseline loss* karena tidak hanya mempertimbangkan luas tumpang tindih (*overlap*) antara *bounding box* prediksi dan *ground truth*, tetapi juga memperhitungkan kesalahan posisi pusat serta perbedaan ukuran *bounding box*. Dengan demikian, kesalahan geometri dapat dikoreksi secara lebih terarah dibandingkan dengan IoU konvensional. Secara matematis, fungsi *loss* EIoU dirumuskan sebagai berikut [3]:

$$L_{EIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{(w - w^{gt})^2}{w_c^2} + \frac{(h - h^{gt})^2}{h_c^2} \quad (1)$$

Untuk meningkatkan fokus pembelajaran pada sampel yang sulit diprediksi, penelitian ini juga mengevaluasi *Focal-EIoU*, yang menambahkan faktor fokus berbasis nilai IoU sehingga prediksi dengan kesalahan tinggi memberikan kontribusi gradien yang lebih besar. Fungsi *loss Focal-EIoU* dirumuskan sebagai berikut [15]: posisi pusat serta perbedaan ukuran *bounding box*. Dengan demikian, kesalahan geometri dapat dikoreksi secara lebih terarah dibandingkan dengan IoU konvensional. Secara matematis, fungsi *loss* EIoU dirumuskan sebagai berikut [3]:

$$L_{FocalEIoU} = (1 - IoU)^\gamma \times L_{EIoU} \quad (2)$$

Selain itu, untuk menangani permasalahan ketidakseimbangan jumlah data antar kelas kematangan, digunakan Class-Balanced Focal-EIoU, yang memberikan bobot berbeda pada setiap kelas berdasarkan distribusi jumlah sampel. Formulasi fungsi *loss* tersebut dituliskan sebagai berikut [18]:

$$L_{CB-FocalEIoU} = \alpha_i \times (1 - IoU)^\gamma \times L_{EIoU} \quad (3)$$

Ketiga fungsi *loss* tersebut diuji secara terpisah dalam kombinasi dengan variasi *feature fusion* dan mekanisme

attention yang telah dijelaskan sebelumnya. Pendekatan ini memungkinkan analisis yang objektif terhadap pengaruh masing-masing fungsi *loss* terhadap akurasi lokalisasi dan konsistensi performa deteksi pada studi ablasi yang dilakukan.

F. Skenario Eksperimen

Eksperimen dirancang menggunakan pendekatan *controlled ablation study* untuk menganalisis pengaruh modifikasi arsitektur dan fungsi *loss* terhadap performa deteksi tingkat kematangan buah mangga berbasis YOLOv10s. Pada setiap eksperimen, hanya komponen tertentu yang divariasikan, sementara komponen lain dipertahankan tetap, sehingga dampak tiap modifikasi dapat dianalisis secara langsung.

Evaluasi difokuskan pada tiga aspek utama, yaitu mekanisme *feature fusion*, mekanisme *attention*, dan fungsi *loss*. Untuk *feature fusion*, digunakan tiga konfigurasi neck, yaitu PAN sebagai baseline, BiFPN dua lapisan, dan BiFPN tiga lapisan, guna menganalisis pengaruh kedalaman fusi fitur terhadap deteksi perbedaan visual yang bersifat halus. Pada aspek *attention*, diuji tiga konfigurasi, yaitu tanpa *attention*, Shuffle Attention, dan Hybrid Attention Transformer, untuk mengevaluasi kontribusi pemodelan konteks lokal dan global.

Seluruh model dilatih menggunakan ukuran input 512×512 piksel dengan parameter pelatihan yang sama pada setiap eksperimen. Dengan pengaturan ini, perbedaan performa yang dihasilkan sepenuhnya mencerminkan pengaruh variasi arsitektur dan fungsi *loss*. Hasil eksperimen kemudian dianalisis menggunakan metrik evaluasi standar untuk menentukan konfigurasi model dengan keseimbangan terbaik antara akurasi deteksi, konsistensi antar kelas, dan efisiensi komputasi.

G. Metode Evaluasi

Evaluasi performa model dilakukan untuk mengukur kemampuan sistem dalam mendeteksi dan mengklasifikasikan tingkat kematangan buah mangga secara akurat. Mengingat permasalahan yang dikaji merupakan deteksi objek multikelas, evaluasi dilakukan menggunakan metrik standar pada tugas deteksi objek, yaitu Precision, Recall, F1-score, dan mean Average Precision (mAP).

Precision digunakan untuk mengukur tingkat ketepatan prediksi model, yaitu proporsi prediksi yang benar terhadap seluruh prediksi yang dihasilkan. Formulasi Precision dinyatakan sebagai berikut [17]:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

Recall digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh objek yang seharusnya terdeteksi. Formulasi Recall dituliskan sebagai berikut [10]:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

Untuk memperoleh ukuran performa yang mempertimbangkan keseimbangan antara *precision* dan *recall*, digunakan metrik *F1-score* yang merupakan rata-rata harmonik dari kedua metrik tersebut. Formulasi *F1-score* dinyatakan sebagai berikut [10]:

$$F1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Selain metrik berbasis klasifikasi, evaluasi performa deteksi objek juga dilakukan menggunakan mean Average

Precision (mAP) untuk mengukur akurasi deteksi secara keseluruhan dengan mempertimbangkan hubungan antara precision dan recall pada berbagai nilai ambang. Secara umum, mAP dirumuskan sebagai berikut [11]:

$$mAP = \frac{1}{N} \sum_{i=1}^N \int_0^1 P_i(R) dR \quad (7)$$

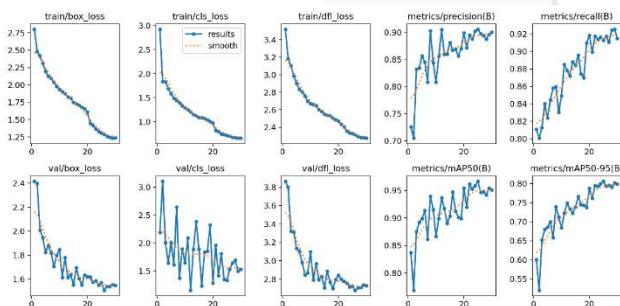
Pada penelitian ini, evaluasi dilakukan menggunakan mAP@0.5, yang menghitung mAP pada ambang Intersection over Union (IoU) sebesar 0.5, serta mAP@0.5:0.95, yang menghitung rata-rata mAP pada rentang ambang IoU dari 0.5 hingga 0.95. Penggunaan kedua metrik ini bertujuan untuk memberikan gambaran yang komprehensif mengenai performa model, baik pada toleransi lokalisasi yang lebih longgar maupun yang lebih ketat.

IV. HASIL DAN PEMBAHASAN

A. Gambaran Umum Hasil Eksperimen

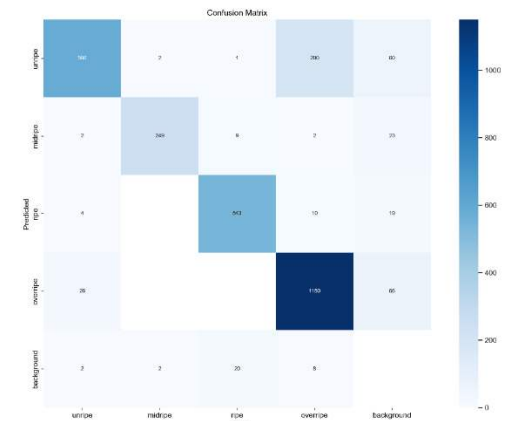
Pada penelitian ini dievaluasi 27 konfigurasi eksperimen yang merupakan kombinasi mekanisme *feature fusion*, *attention*, dan fungsi *loss* untuk mengidentifikasi arsitektur terbaik dalam deteksi tingkat kematangan buah mangga multikelas pada lingkungan kebun alami. Hasil eksperimen menunjukkan bahwa kombinasi BiFPN tiga lapisan (BiFPN(3)), Hybrid Attention Transformer (HAT), dan fungsi loss EIou memberikan performa terbaik dengan nilai mAP tinggi serta keseimbangan precision dan recall yang baik. Hal ini menunjukkan bahwa fusi fitur multi-skala yang lebih dalam dan pemodelan konteks global berperan penting dalam membedakan perbedaan visual kematangan yang bersifat halus.

Perilaku pelatihan model terbaik ditunjukkan melalui kurva training dan validation pada Gambar 7. Kurva loss menunjukkan penurunan yang konsisten seiring bertambahnya jumlah epoch, sementara metrik evaluasi meningkat secara bertahap hingga mencapai kondisi konvergen. Kesesuaian performa antara data pelatihan dan validasi menunjukkan bahwa model memiliki stabilitas pelatihan yang baik dan tidak mengalami overfitting yang signifikan.



GAMBAR 7
Training dan Kurva Model Terbaik

Analisis confusion matrix pada Gambar 8 menunjukkan bahwa kelas overripe memiliki tingkat prediksi benar tertinggi dibandingkan kelas kematangan lainnya. Hal ini berkaitan dengan karakteristik visual buah overripe yang relatif lebih kontras. Sebaliknya, kesalahan klasifikasi paling sering terjadi antara kelas mid-ripe dan ripe, yang memiliki perbedaan warna dan tekstur yang sangat halus serta sensitif terhadap variasi pencahayaan.



GAMBAR 8
Confusion Matrix Model Terbaik

Contoh hasil deteksi tingkat kematangan buah mangga menggunakan model terbaik ditunjukkan pada Gambar 4.1.3. Pada visualisasi tersebut, model berhasil melokalisasi objek buah mangga serta memprediksi kelas tingkat kematangan dengan nilai confidence yang tinggi. Hasil ini menunjukkan bahwa model tidak hanya akurat secara kuantitatif, tetapi juga konsisten secara visual pada berbagai kondisi citra pengujian.



GAMBAR 9
Contoh hasil deteksi tingkat kematangan buah mangga

Secara keseluruhan, hasil pada subbab ini menunjukkan bahwa pendekatan yang diusulkan mampu meningkatkan performa deteksi tingkat kematangan buah mangga secara signifikan. Temuan ini menjadi dasar untuk analisis lebih lanjut mengenai pengaruh masing-masing komponen arsitektur, yang akan dibahas pada subbab berikutnya.

B. Pengaruh Modul Feature Fusion

Subbab ini mengevaluasi pengaruh mekanisme feature fusion terhadap performa deteksi tingkat kematangan buah mangga dengan membandingkan tiga konfigurasi neck, yaitu Path Aggregation Network (PAN), Bidirectional Feature Pyramid Network dengan dua lapisan (BiFPN(2)), dan Bidirectional Feature Pyramid Network dengan tiga lapisan (BiFPN(3)).

Hasil perbandingan performa ketiga konfigurasi ditunjukkan pada Tabel 3. Secara umum, kedua konfigurasi BiFPN menunjukkan peningkatan performa yang konsisten dibandingkan PAN pada seluruh metrik evaluasi. Di antara

ketiganya, BiFPN(3) menghasilkan performa terbaik dengan nilai mAP dan F1-score tertinggi.

TABEL 3

RATA-RATA PERFORMA FEATURE FUSION DARI 27 EKSPERIMEN.

Fusion	mAP@0.5	mAP@0.5:0.95	R	P	F1
PAN	0.949	0.793	0.929	0.909	0.911
BiFPN (2)	0.953	0.804	0.931	0.912	0.912
BiFPN (3)	0.957	0.806	0.934	0.913	0.913

Peningkatan performa pada BiFPN menunjukkan bahwa fusi fitur dua arah dengan bobot adaptif mampu menggabungkan informasi visual lintas skala secara lebih efektif. Konfigurasi BiFPN(3) memberikan peningkatan paling konsisten karena memungkinkan interaksi fitur yang lebih kaya, sehingga membantu model membedakan tingkat kematangan buah dengan karakteristik visual yang mirip. Meskipun memiliki kompleksitas komputasi yang sedikit lebih tinggi, peningkatan akurasi yang diperoleh masih berada pada tingkat yang dapat diterima.

C. Pengaruh Mekanisme Attention

Subbab ini mengevaluasi pengaruh mekanisme attention terhadap performa model dalam mendeteksi tingkat kematangan buah mangga. Evaluasi dilakukan dengan membandingkan tiga konfigurasi, yaitu tanpa attention, penggunaan Shuffle Attention (SA), dan Hybrid Attention Transformer (HAT).

Hasil perbandingan performa ketiga konfigurasi ditunjukkan pada Tabel 4. Secara umum, penerapan mekanisme attention meningkatkan performa deteksi dibandingkan konfigurasi tanpa attention. Shuffle Attention memberikan peningkatan moderat pada seluruh metrik evaluasi, sedangkan Hybrid Attention Transformer menghasilkan performa terbaik, khususnya pada nilai mAP dan F1-score.

TABEL 4

RATA-RATA PERFORMA MEKANISME ATTENTION DARI 27 EKSPERIMEN.

Attention	mAP@0.5	mAP@0.5:0.95	R	P	F1
None	0.950	0.797	0.926	0.904	0.906
Shuffle Attention	0.955	0.803	0.930	0.908	0.909
Hybrid Attention Transformer	0.961	0.806	0.934	0.911	0.912

Peningkatan pada Shuffle Attention menunjukkan bahwa perhatian kanal dan spasial membantu menonjolkan fitur visual penting, namun masih terbatas karena fokus pada pemrosesan lokal. Sebaliknya, HAT mampu memodelkan hubungan global antar fitur melalui *self-attention*, sehingga lebih efektif dalam membedakan kelas dengan perbedaan visual yang sangat halus, seperti mid-ripe dan ripe. Secara keseluruhan, hasil ini menunjukkan bahwa mekanisme *attention*, khususnya Hybrid Attention Transformer, berperan penting dalam meningkatkan kemampuan model menangkap perbedaan visual halus dan menjadi komponen pendukung utama dalam arsitektur yang diusulkan.

D. Pengaruh Fungsi Loss

Subbab ini mengevaluasi pengaruh pemilihan fungsi loss terhadap performa deteksi tingkat kematangan buah mangga. Evaluasi dilakukan dengan membandingkan tiga fungsi loss,

yaitu Efficient Intersection over Union (EIoU), Focal-EIoU, dan Class-Balanced Focal-EIoU, dengan konfigurasi arsitektur lainnya dipertahankan tetap agar pengaruh fungsi loss dapat dianalisis secara objektif.

Hasil perbandingan performa ketiga fungsi loss ditunjukkan pada Tabel 5. Secara umum, EIoU menghasilkan performa paling seimbang pada seluruh metrik evaluasi, termasuk mAP, precision, recall, dan F1-score. Hal ini menunjukkan bahwa formulasi EIoU mampu menghasilkan prediksi bounding box yang lebih stabil dan konsisten.

TABEL 5

RATA-RATA PERFORMA FUNGSI LOSS DARI 27 EKSPERIMEN.

Loss	mAP@0.5	mAP@0.5:0.95	R	P	F1
EIoU	0.957	0.806	0.934	0.913	0.910
Focal-EIoU	0.955	0.803	0.938	0.909	0.909
Class-Balanced Focal-EIoU	0.954	0.802	0.940	0.905	0.907

Focal-EIoU dan Class-Balanced Focal-EIoU menunjukkan peningkatan nilai recall, menandakan sensitivitas yang lebih tinggi terhadap sampel sulit. Namun, peningkatan tersebut diikuti oleh penurunan precision akibat bertambahnya *false positive*, terutama pada kelas dengan kemiripan visual tinggi. Hasil ini menunjukkan bahwa pada dataset dengan perbedaan visual yang bersifat gradual, pendekatan *loss* yang terlalu agresif tidak selalu optimal. Oleh karena itu, EIoU dipilih sebagai fungsi *loss* pada konfigurasi akhir karena memberikan keseimbangan terbaik antara akurasi lokalisasi dan stabilitas prediksi.

V. KESIMPULAN

Berdasarkan hasil penelitian, pengembangan model deteksi tingkat kematangan buah mangga berbasis YOLOv10s terbukti mampu meningkatkan performa deteksi secara signifikan pada skenario multikelas dengan perbedaan visual yang halus. Penggunaan feature fusion multi-skala melalui BiFPN(3) meningkatkan kualitas representasi fitur, sementara integrasi Hybrid Attention Transformer membantu model memfokuskan perhatian pada fitur visual yang relevan. Dari sisi fungsi loss, EIoU memberikan keseimbangan terbaik antara akurasi lokalisasi dan stabilitas prediksi pada dataset yang memiliki karakteristik visual gradual. Secara keseluruhan, hasil penelitian menunjukkan bahwa optimisasi arsitektur YOLO melalui studi ablasi sistematis efektif meningkatkan akurasi deteksi tanpa mengorbankan efisiensi komputasi, sehingga sesuai untuk penerapan deteksi kematangan buah mangga secara real-time di lingkungan pertanian.

REFERENSI

- [1] I. Esaki, S. Noma, T. Ban, R. Sultana, and I. Shimizu, "Maturity Classification of Blueberry Fruit Using YOLO and Vision Transformer for Agricultural Assistance," *Sensors*, vol. 23, no. 14, p. 6395, 2023.
- [2] R. Li, Y. Li, W. Qin, A. Abbas, S. Li, R. Ji, Y. Wu, Y. He, and J. Yang, "Lightweight Network for Corn Leaf Disease Identification Based on Improved YOLOv8s," *Computers and Electronics in Agriculture*, vol. 217, p. 108687, 2024.

- [3] J. Tang, Z. Yu, and C. Shao, "Hybrid Attention Transformer Integrated YOLOv8 for Fruit Ripeness Detection," *Scientific Reports*, vol. 15, p. 22652, 2025.
- [4] Y. Wang, C. Ouyang, H. Peng, J. Deng, L. Yang, H. Chen, Y. Luo, and P. Jiang, "YOLO-ALW: An Enhanced High-Precision Model for Chili Maturity Detection," *Agriculture*, vol. 15, p. 1400, 2025.
- [5] J. J. K. Chai, J.-L. Xu, and C. O'Sullivan, "Real-Time Detection of Strawberry Ripeness Using Augmented Reality and Deep Learning," *Sensors*, vol. 23, no. 17, p. 7639, 2023.
- [6] W. Huang, Y. Liao, P. Wang, Z. Chen, Z. Yang, L. Xu, and J. Mu, "AITP-YOLO: Improved Tomato Ripeness Detection Model Based on Multiple Strategies," *Frontiers in Plant Science*, vol. 16, p. 1596739, 2025.
- [7] J. Ren, W. Wang, Y. Tian, and J. He, "YOLOv10-Kiwi: A YOLOv10-Based Lightweight Kiwifruit Detection Model in Trellised Orchards," *Frontiers in Plant Science*, vol. 16, p. 1616165, 2025.
- [8] L. Wang, R. Ye, Y. Chen, and T. Li, "YOLOv10-LGDA: An Improved Algorithm for Defect Detection in Citrus Fruits Across Diverse Backgrounds," *Plants*, vol. 14, p. 1990, 2025.
- [9] H. Fu, Z. Guo, Q. Feng, F. Xie, Y. Zuo, and T. Li, "MSOAR-YOLOv10: Multi-Scale Occluded Apple Detection for Enhanced Harvest Robotics," *Horticulturae*, vol. 10, p. 1246, 2024.
- [10] P. Kamat, S. Gite, H. Chandekar, L. Dlima, and B. Pradhan, "Multi-class Fruit Ripeness Detection Using YOLO and SSD," *Journal of Food Engineering and Automation*, vol. 9, no. 3, pp. 112--124, 2022.
- [11] A. A. Luthfie and A. Alamsyah, "Identifying Coconut Maturity Levels Using CNN and YOLOv8," *Agriculture*, vol. 14, no. 4, p. 812, 2024.
- [12] Y. Xu, H. Li, Y. Zhou, Y. Zhai, Y. Yang, and D. Fu, "GLL-YOLO: A Lightweight Network for Detecting the Maturity of Blueberry Fruits," *Agriculture*, vol. 15, p. 1877, 2025.
- [13] Z. Huang, X. Li, S. Fan, Y. Liu, H. Zou, X. He, S. Xu, J. Zhao, and W. Li, "ORD-YOLO: A Ripeness Recognition Method for Citrus Fruits in Complex Environments," *Agriculture*, vol. 15, p. 1711, 2025.
- [14] B. Ma, Z. Hua, Y. Wen, H. Deng, Y. Zhao, L. Pu, and H. Song, "Using an Improved Lightweight YOLOv8 Model for Real-Time Detection of Multi-Stage Apple Fruit," *Artificial Intelligence in Agriculture*, vol. 11, pp. 70--82, 2024.
- [15] C. Wang, H. Wang, Q. Han, Z. Zhang, D. Kong, and X. Zou, "Strawberry Detection and Ripeness Classification Using YOLOv8+," *Agriculture*, vol. 14, p. 751, 2024.
- [16] F. Yuan, J. Wang, W. Ding, S. Mei, C. Fang, S. Chen, and H. Zhou, "A Lightweight and Rapid Dragon Fruit Detection Method," *Agriculture*, vol. 15, p. 1120, 2025.
- [17] M. Nergiz, "Enhancing Strawberry Harvesting Efficiency through YOLO-v7 Object Detection Assessment," *Turkish Journal of Science and Technology*, vol. 18, pp. 519-533, 2023.
- [18] H. C. Bazame, J. P. Molin, D. Althoff, and M. Martello, "Detection of Coffee Fruits on Tree Branches Using Computer Vision," *Scientia Agricola*, vol. 80, p. e20220064, 2023.
- [19] R. E. Saragih, A. Purnajaya, I. Syafrinal, Y. Fernando, and Yodi, "Mango and Banana Ripeness Detection Based on Lightweight YOLOv8," *Jurnal Buana Informatika*, vol. 15, no. 2, pp. 79--88, 2024.
- [20] A. Aljaafreh, E. Elzagzoug, J. Abukhait, A.-H. Soliman, S. Alja'afreh, A. Sivanathan, and J. Hughes, "A Real-Time Olive Fruit Detection for Harvesting Robot Based on YOLO Algorithms," *Journal of Robotics and Control*, vol. 4, no. 2, pp. 111--121, 2023.
- [21] X. Gao, J. Du, X. Liu, D. Jia, and J. Wang, "Object Detection Based on Improved YOLOv10 for Electrical Equipment," *Processes*, vol. 13, no. 529, 2025.
- [22] B. Liu, X. Wei, Q. Chen, J. Liu, Y. Chen, P. Yu, S. Lei, and Y. Hu, "Safety Helmet Detection Methods in Heavy Machinery Factory," 2024.
- [23] F. Zhincheng, Y. Jie, C. Lifang, C. Zhichao, L. Linhong, "An Intelligent Waste-Sorting and Recycling Device Based on Improved EfficientNet", *International Journal of Environmental Research and Public Health (IJERPH)*, vol 19(23):15987.
- [24] X. Yang, J. Zhao, L. Zhao, I. Ganchev, et al., "Detection of River Floating Garbage Based on Improved YOLOv5," *Mathematics*, vol. 10, no. 22, p. 4366, 2022, doi: 10.3390/math10224366.
- [25] H. T. Nguyen, T. Le, T. V. Nguyen, and N. T. Nguyen, "Enhancing MRI Brain Tumor Segmentation with an Additional Classification Network," 2020, doi: 10.48550/arXiv.2009.12111.
- [26] Z. Wang, S. Wan, and X. Ma, "Remote Sensing Image Dense Target Detection Based on Rotating Frame," *Journal of Physics: Conference Series*, vol. 2006, no. 1, p. 012049, 2021, doi: 10.1088/1742-6596/2006/1/012049.