TESIS

ANALISIS EKSPERIMENTAL PENGGANTIAN *BACKBONE* YOLOv5 DENGAN EFFICIENTNET-B4 PADA SISTEM DETEKSI NILAI MATA UANG

disusun oleh

ALFRIDA SABAR

23.55.2514

Konsentrasi : Digital Transformation Intelligence

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

TESIS

ANALISIS EKSPERIMENTAL PENGGANTIAN *BACKBONE* YOLOv5 DENGAN EFFICIENTNET-B4 PADA SISTEM DETEKSI NILAI MATA UANG

EXPERIMENTAL ANALYSIS OF REPLACING THE YOLOV5 *BACKBONE* WITH EFFICIENTNET-B4 IN A CURRENCY VALUE DETECTION SYSTEM

Diajukan untuk memenuhi salah satu syarat mencapai derajat Pascasarjana
Program Studi *(PJJ Informatika)*

disusun oleh

ALFRIDA SABAR

23.55.2514

Konsentrasi : Digital Transformation Intelligence

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

HALAMAN PERSETUJUAN

ANALISIS EKSPERIMENTAL PENGGANTIAN *BACKBONE* YOLOv5 DENGAN EFFICIENTNET-B4 PADA SISTEM DETEKSI NILAI MATA UANG

EXPERIMENTAL ANALYSIS OF REPLACING THE YOLOV5 *BACKBONE* WITH EFFICIENTNET-B4 IN A CURRENCY VALUE DETECTION SYSTEM

yang disusun dan diajukan oleh

Alfrida Sabar

23.55.2514

telah disetujui oleh Dosen Pembimbing Tesis
pada tanggal 5 Juni 2026

Dosen Pembimbing,

Prof. Dr. Ema Utami, S.Si., M.Kom.
NIK. 190302037

HALAMAN PENGESAHAN

ANALISIS EKSPERIMENTAL PENGGANTIAN *BACKBONE* YOLOv5 DENGAN EFFICIENTNET-B4 PADA SISTEM DETEKSI NILAI MATA UANG

EXPERIMENTAL ANALYSIS OF REPLACING THE YOLOV5 *BACKBONE* WITH EFFICIENTNET-B4 IN A CURRENCY VALUE DETECTION SYSTEM

yang disusun dan diajukan oleh

Alfrida Sabar

23.55.2514

Telah dipertahankan di depan Dewan Penguji
pada tanggal 5 Juni 2026

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Hanif Al Fatta, S.Kom., M.Kom., Ph.D. **NIK. 190302096**	
Dr. Andi Sunyoto, S.Kom., M.Kom. **NIK. 190302052**	
Prof. Dr. Ema Utami, S.Si., M.Kom. **NIK. 190302037**	

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer
Tanggal 5 Juni 2026

DEKAN FAKULTAS ILMU KOMPUTER

Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Alfrida Sabar
NIM : 23.55.2514

Menyatakan bahwa Tesis dengan judul berikut:

Analisis Eksperimental Penggantian *Backbone* YOLOv5 Dengan EfficientNet-B4 Pada Sistem Deteksi Nilai Mata Uang

Dosen Pembimbing : Prof. Dr. Ema Utami, S.Si., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian saya sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab saya, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini saya buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 5 Juni 2026

Yang Menyatakan,

Alfrida Sabar

HALAMAN PERSEMBAHAN

Puji Syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang senantiasa memberikan rahmat dan kesehatan, sehingga penulis dapat menyelesaikan Thesis ini sebagai salah satu syarat untuk mendapatkan gelar Magister Komputer. Walaupun Thesis ini jauh dari kata sempurna, namun penulis bangga telah mencapai titik ini dimana Thesis ini dapat selesai di waktu yang tepat.

Thesis ini penulis persembahkan kepada:

1. Orang tua terkasih yang telah memberikan dukungan moril serta doa yang tiada henti untuk kesuksesan penulis.
2. Suami tercinta yang selalu mendampingi dan memberikan dukungan tanpa henti dalam setiap kondisi, baik suka maupun duka, selama proses penyusunan tesis ini.
3. Saudara-saudara dan keluarga besar yang senantiasa mendoakan serta memberi dorongan penuh kepada penulis.
4. Teman dan sahabat-sahabat terbaik yang setia menemani, membantu, memberi motivasi, serta menjadi bagian dari perjalanan panjang penulis.
5. Bapak dan Ibu Dosen pembimbing, penguji dan pengajar yang selama ini telah tulus dan ikhlas meluangkan waktunya untuk menuntun dan mengarahkan spenulis, memberikan bimbingan dan pelajaran yang tiada ternilai harganya.
6. Kepada semua pihak yang telah membantu, terima kasih atas kontribusi dan waktunya.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena berkat rahmat dan karunia-Nya, tesis ini dapat diselesaikan dengan baik. Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister Komputer pada Jurusan Informatika di Universitas Amikom Yogyakarta.

Dalam proses penyusunan tesis ini, penulis telah banyak menerima bimbingan, arahan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Prof. Dr. Ema Utami, S.Si., M.Kom selaku Dosen Pembimbing I dan Bapak Dr. Ferry Wahyu Wibowo, S.Si., M.Cs selaku Dosen Pembimbing II, yang telah membimbing penulis dari awal hingga akhir penyusunan tesis.
2. Pimpinan serta rekan-rekan Kantor Loka Monitor SFR Tanjung Selor atas dukungan selama menjalani pendidikan.
3. Teman-teman seperjuangan yang selalu memberi semangat dalam menyelesaikan tesis ini.
4. Semua pihak yang telah membantu dalam proses penelitian ini.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa yang akan datang.

Akhir kata, penulis berharap semoga tesis ini dapat memberikan manfaat bagi semua pihak yang membutuhkan.

Yogyakarta, 5 Juni 2026
Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR ISTILAH

CNN	*Convolutional Neural Network*
FPN	Feature Pyramid Network
PANet	Path Aggregation Network
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
mAP	mean Average Precision

INTISARI

YOLOv5 merupakan salah satu varian algoritma *You Only Look Once* (YOLO) yang dikenal mampu melakukan deteksi objek secara cepat dan efisien dalam satu tahap pemrosesan. Kinerja YOLOv5 sangat dipengaruhi oleh arsitektur *backbone* yang berperan dalam mengekstraksi fitur visual dari citra input. Oleh karena itu, penggantian *backbone* menjadi salah satu pendekatan yang digunakan untuk mengevaluasi pengaruh kualitas ekstraksi fitur terhadap performa deteksi objek.

Penelitian ini bertujuan untuk melakukan evaluasi eksperimental penggantian *backbone* YOLOv5 menggunakan arsitektur EfficientNet, khususnya EfficientNet-B4, pada sistem deteksi nilai mata uang rupiah. Dataset yang digunakan terdiri dari delapan kelas, yaitu tujuh kelas nominal uang rupiah dan satu kelas non-uang. Eksperimen dilakukan dengan membandingkan performa YOLOv5 baseline dan YOLOv5 dengan *backbone* EfficientNet-B4 menggunakan strategi *transfer learning* dan *freeze–unfreeze*. Evaluasi model dilakukan menggunakan metrik precision, recall, mAP@0.5, mAP@0.5:0.95, serta waktu *inference*.

Hasil pengujian menunjukkan bahwa penggunaan EfficientNet-B4 meningkatkan nilai precision dari 74.6% menjadi 77.4%, recall dari 59.6% menjadi 84.7%, mAP@0.5 dari 69.5% menjadi 90.1%, serta mAP@0.5:0.95 dari 48.0% menjadi 74.6%. Peningkatan tersebut menunjukkan bahwa EfficientNet-B4 mampu meningkatkan kemampuan deteksi objek sekaligus kualitas lokalisasi *bounding box* pada berbagai threshold IoU. Dari sisi komputasi, waktu *inference* meningkat dari 6.3 ms menjadi 40.8 ms per gambar akibat kompleksitas model yang lebih besar. Selain itu, perbandingan EfficientNet-B0, B1, dan B4 menunjukkan bahwa peningkatan kompleksitas backbone tidak selalu menghasilkan performa deteksi yang lebih tinggi. EfficientNet-B0 memperoleh nilai mAP@0.5 sebesar 91.9% dan mAP@0.5:0.95 sebesar 77.7%, lebih tinggi dibandingkan EfficientNet-B1 dan EfficientNet-B4, serta memiliki waktu *inference* yang lebih cepat. Hasil penelitian

ini menunjukkan bahwa penggantian backbone YOLOv5 dengan EfficientNet-B4 mampu meningkatkan kualitas deteksi nilai mata uang rupiah dibandingkan YOLOv5 baseline, khususnya pada aspek kemampuan deteksi objek dan kualitas lokalisasi *bounding box*.

Kata kunci: YOLOv5, EfficientNet-B4, Currency Detection, Currency Recognition.

ABSTRACT

YOLOv5 is a variant of the You Only Look Once (YOLO) algorithm known for its ability to detect objects quickly and efficiently in a single processing step. YOLOv5's performance is heavily influenced by its backbone architecture, which is responsible for extracting visual features from the input images. Therefore, replacing the backbone is one approach used to evaluate the impact of feature extraction quality on object detection performance.

This study aims to conduct an experimental evaluation of replacing the YOLOv5 backbone with the EfficientNet architecture, specifically EfficientNet-B4, in a system for detecting Indonesian rupiah banknotes. The dataset used consists of eight classes: seven classes of rupiah denominations and one non-currency class. The experiment was conducted by comparing the performance of the YOLOv5 baseline and YOLOv5 with the EfficientNet-B4 backbone using transfer learning and freeze–unfreeze strategies. Model evaluation was performed using the metrics precision, recall, mAP@0.5, mAP@0.5:0.95, and inference time.

Test results show that using EfficientNet-B4 improves precision from 74.6% to 77.4%, recall from 59.6% to 84.7%, the mAP@0.5 from 69.5% to 90.1%, and the mAP@0.5:0.95 from 48.0% to 74.6%. These improvements indicate that EfficientNet-B4 is capable of enhancing both object detection performance and the quality of bounding box localization across various IoU thresholds. In terms of computational efficiency, inference time increased from 6.3 ms to 40.8 ms per image due to the model's greater computational complexity. Furthermore, a comparison of EfficientNet-B0, B1, and B4 shows that increasing the backbone's complexity does not always result in higher detection performance. EfficientNet-B0 achieved a mAP@0.5 of 91.9% and a mAP@0.5:0.95 of 77.7%, which are higher than those of EfficientNet-B1 and EfficientNet-B4, and it also has a faster inference time. The results of this study indicate that replacing the YOLOv5 backbone with EfficientNet-B4 improves the detection quality of the Indonesian rupiah compared

to the YOLOv5 baseline, particularly in terms of object detection capability and bounding box localization quality.

Keyword: *YOLOv5, EfficientNet-B4, Currency Detection, Currency Recognition.*

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi Artificial Intelligence (AI) telah membawa perubahan besar dalam cara sistem pengenalan dan deteksi objek bekerja, termasuk deteksi nilai mata uang. AI memungkinkan komputer untuk memahami pola-pola yang sulit dikenali, seperti perbedaan halus pada angka, warna, atau desain pada uang kertas. Teknologi ini juga mampu beradaptasi dengan berbagai kondisi, seperti cahaya yang redup, sudut pandang yang berbeda, atau kondisi uang kertas yang telah usang. Sistem berbasis AI mampu mengatasi tantangan dalam pengenalan uang kertas, termasuk dalam berbagai kondisi fisik seperti lipatan, rusak serta di bawah pencahayaan yang beragam, sehingga memastikan akurasi dan keandalan yang tinggi dalam penggunaannya [1].

Saat ini banyak penelitian telah mengembangkan cara-cara mendeteksi objek dan banyak dari penelitian tersebut merekomendasikan penggunaan algoritma YOLO untuk deteksi objek secara langsung. Algoritma YOLO dianggap memiliki akurasi deteksi yang tinggi dan kecepatan yang lebih baik dibandingkan metode lain, sehingga cocok untuk mendeteksi objek secara real-time. Algoritma ini pertama kali diperkenalkan oleh Joseph Redmon et al, dalam penelitiannya menemukan metode untuk mendeteksi objek secara langsung [2]. Dalam penelitian tersebut, YOLO terbukti mampu mendeteksi objek dengan kecepatan dan akurasi yang sangat tinggi. YOLO memproses seluruh gambar hanya dalam satu kali proses

atau sekali umpan (*single shot*) untuk mendeteksi semua objek sekaligus, dan mampu memberikan hasil deteksi objek secara real-time.

Asitektur YOLO terdiri dari tiga komponen utama yaitu *backbone*, *neck*, dan *head* yang masing-masing memiliki fungsi berbeda. Backbone merupakan bagian penting yang berfungsi mengekstrak fitur visual dari gambar input dan mengubahnya menjadi representasi fitur yang lebih abstrak [3]. Dalam sistem *object detection*, tantangan utama tidak hanya terletak pada kemampuan model dalam mengenali kelas objek, tetapi juga pada kemampuan model dalam melakukan lokalisasi objek secara akurat. Diwan et al menjelaskan bahwa *localization error* masih menjadi salah satu permasalahan utama pada algoritma YOLO, terutama pada objek dengan karakteristik visual yang mirip dan kondisi latar belakang yang kompleks [4]. Kesalahan lokalisasi dapat menyebabkan prediksi *bounding box* tidak sesuai dengan posisi objek sebenarnya sehingga menurunkan kualitas deteksi objek. Permasalahan ini menjadi lebih kompleks pada deteksi uang rupiah karena adanya kemiripan antar pecahan, lipatan uang, perubahan sudut pengambilan gambar, dan variasi pencahayaan. Begitupun penelitian Ramos and Sappa berjudul "*A Decade of YOLO for Object Detection*" menjelaskan bahwa variasi skala objek, *occlusion*, perubahan posisi objek, serta *intra-class variation* masih menjadi tantangan utama dalam sistem object detection modern [5]. Pada kasus uang rupiah, kemiripan warna dan ornamen antar pecahan uang menyebabkan model tidak hanya dituntut mampu mengenali kelas objek, tetapi juga harus mampu melakukan lokalisasi objek secara lebih presisi.

Selain itu, penelitian Ma et al menyebutkan bahwa *localization error* merupakan salah satu faktor utama yang membatasi performa sistem *object detection* [6]. Kondisi ini menunjukkan bahwa kualitas ekstraksi fitur dan ketepatan penentuan posisi objek menjadi aspek penting dalam pengembangan sistem deteksi objek modern. Pada arsitektur YOLO, kualitas ekstraksi fitur sangat dipengaruhi oleh backbone yang digunakan. Backbone berperan sebagai *feature extractor* yang bertugas menghasilkan representasi visual objek sebelum diteruskan ke bagian *neck* dan *head* untuk proses deteksi. Jika fitur yang dihasilkan kurang representatif, maka model akan lebih sulit membedakan karakteristik antar objek dan menentukan posisi objek secara akurat. Akibatnya, *localization error* dapat meningkat dan kualitas *bounding box* yang dihasilkan menjadi kurang presisi. Oleh karena itu, peningkatan kualitas ekstraksi fitur pada backbone menjadi salah satu pendekatan penting untuk meningkatkan performa deteksi objek dan mengurangi kesalahan lokalisasi.

Beberapa penelitian sebelumnya telah melakukan penggantian backbone pada arsitektur YOLO menggunakan berbagai jaringan konvolusional, seperti EfficientNet, MobileNet, ResNet, dan DenseNet untuk meningkatkan performa deteksi objek. Meena et al. menjelaskan bahwa pendekatan *hybrid neural network* mampu meningkatkan performa pengenalan objek dibandingkan metode tradisional [7]. Dalam penelitian ini digunakan arsitektur EfficientNet-B4 sebagai backbone alternatif pada YOLOv5 karena memiliki kemampuan ekstraksi fitur yang baik melalui prinsip *compound scaling*. Penelitian ini bertujuan untuk menganalisis pengaruh penggunaan EfficientNet-B4 sebagai backbone YOLOv5 terhadap

peningkatan kualitas ekstraksi fitur dan ketepatan lokalisasi objek pada sistem deteksi nilai mata uang rupiah, serta dampaknya terhadap efisiensi komputasi model.

Meskipun beberapa penelitian telah menunjukkan bahwa penggantian *backbone* dapat meningkatkan performa deteksi objek, kajian yang secara khusus mengevaluasi penggunaan EfficientNet-B4 sebagai *backbone* pengganti CSPDarknet pada YOLOv5 untuk meningkatkan kualitas ekstraksi fitur dan mengurangi *localization error* pada deteksi nilai mata uang rupiah masih terbatas. Selain itu, penelitian terdahulu belum banyak membahas pengaruh penggunaan EfficientNet-B4 terhadap ketepatan lokalisasi objek yang diukur menggunakan metrik mAP@0.5 dan mAP@0.5:0.95. Oleh karena itu, penelitian ini mengusulkan penggunaan EfficientNet-B4 sebagai backbone alternatif pada YOLOv5 untuk meningkatkan kualitas representasi fitur, memperbaiki presisi bounding box, serta meningkatkan performa deteksi nilai mata uang rupiah.

Berdasarkan *research gap* tersebut, kontribusi utama penelitian ini adalah mengusulkan penggunaan EfficientNet-B4 sebagai backbone pengganti CSPDarknet pada YOLOv5 untuk meningkatkan kualitas ekstraksi fitur dan ketepatan lokalisasi objek pada sistem deteksi nilai mata uang rupiah. Kebaruan penelitian ini terletak pada penerapan EfficientNet-B4 sebagai backbone alternatif YOLOv5 yang secara khusus dievaluasi terhadap kemampuan lokalisasi objek menggunakan metrik mAP@0.5 dan mAP@0.5:0.95. Selain itu, penelitian ini menerapkan strategi pelatihan bertahap (*freeze–unfreeze training strategy*), yaitu membekukan bobot backbone pada tahap awal pelatihan untuk mempertahankan

pengetahuan hasil pretraining, kemudian membuka kembali seluruh layer untuk proses fine-tuning terhadap karakteristik dataset uang rupiah. Strategi tersebut digunakan untuk mengoptimalkan proses transfer learning dan meningkatkan kualitas representasi fitur yang dihasilkan model. Penelitian ini juga memberikan analisis trade-off antara peningkatan performa deteksi dan efisiensi komputasi melalui pengukuran waktu inferensi. Hasil penelitian menunjukkan bahwa penggunaan EfficientNet-B4 yang dikombinasikan dengan strategi freeze–unfreeze mampu meningkatkan Recall, mAP@0.5, dan mAP@0.5:0.95 dibandingkan YOLOv5 baseline, sehingga memberikan kontribusi empiris mengenai pengaruh kualitas backbone dan strategi pelatihan terhadap performa deteksi nominal uang rupiah.

Ke depannya, sistem ini diharapkan dapat dikembangkan dan diterapkan pada berbagai perangkat, seperti alat bantu deteksi nilai mata uang untuk penyandang tuna netra, mesin kasir otomatis, mesin teller atau ATM, layanan penukaran uang (*money changer*) untuk wisatawan, serta alat deteksi uang palsu. Dengan begitu, sistem ini dapat membantu mempermudah transaksi keuangan, meningkatkan keamanan, mendukung berbagai kebutuhan masyarakat, dan membantu mengurangi peredaran uang palsu.

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian, permasalahan utama yang dikaji dalam penelitian ini dirumuskan sebagai berikut:

1. Bagaimana kinerja model YOLOv5 dalam mendeteksi nilai mata uang kertas rupiah?

2. Bagaimana kinerja model YOLOv5 dengan *backbone* EfficientNet-B4 dalam mendeteksi nilai mata uang kertas rupiah?
3. Bagaimana pengaruh penggantian *backbone* YOLOv5 (CSPDarknet) dengan *backbone* EfficientNet-B4 terhadap kinerja sistem deteksi nilai mata uang rupiah?
4. Bagaimana perbandingan kinerja YOLOv5 dengan *backbone* EfficientNet-B4 terhadap EfficientNet-B0 dan EfficientNet-B1 dalam sistem deteksi nilai mata uang rupiah?

1.3 Batasan Masalah

Agar penelitian ini terfokus dan memiliki batas kajian yang jelas, maka batasan masalah dalam penelitian ini ditetapkan sebagai berikut:

a. Objek penelitian dibatasi pada sistem deteksi nilai mata uang kertas rupiah berbasis citra digital menggunakan pendekatan object detection, bukan sistem klasifikasi keaslian uang, deteksi uang palsu, maupun sistem berbasis video secara *real-time*.
b. Dataset yang digunakan dibatasi pada citra uang kertas rupiah yang masih berlaku dan beredar secara resmi di Indonesia saat ini dari berbagai nominal yang resmi diterbitkan oleh Bank Indonesia.
c. Model deteksi objek yang dianalisis dibatasi pada arsitektur YOLOv5 dengan tiga konfigurasi *backbone*, yaitu:
    1. YOLOv5 baseline dengan *backbone* CSPDarknet,
    2. YOLOv5 dengan *backbone* EfficientNet-B4,

3. YOLOv5 dengan *backbone* EfficientNet-B0 dan EfficientNet-B1 sebagai pembanding tambahan.

d. Modifikasi arsitektur hanya dilakukan pada bagian *backbone* , tanpa mengubah struktur neck dan head deteksi YOLOv5.

e. Strategi pelatihan model dibatasi pada konfigurasi berikut:

- YOLOv5 baseline dilatih end-to-end selama 30 epoch menggunakan optimizer SGD.
- YOLOv5 dengan *backbone* EfficientNet-B4 dilatih menggunakan pendekatan dua tahap (Stage A – Freeze 30 epoch, Stage B – Unfreeze 30 epoch) dengan optimizer AdamW.
- Resolusi input 640×640 untuk seluruh model.
- Pelatihan dilakukan menggunakan GPU (CUDA) pada Google Colab.
- Parameter batch size disesuaikan dengan kompleksitas *backbone* (16 untuk YOLOv5 baseline dan 4 untuk YOLOv5 dengan *backbone* EfficientNet).

f. Evaluasi performa model dibatasi pada metrik kuantitatif yaitu Precision, Recall, mean Average Precision (mAP@0.5 dan mAP@0.5:0.95), inference time per citra dan *percentage improvement* terhadap YOLOv5 baseline.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Menganalisis kinerja model YOLOv5 baseline dalam mendeteksi nilai mata uang kertas rupiah pada berbagai kondisi visual.

2. Mengimplementasikan dan mengevaluasi model YOLOv5 dengan *backbone* EfficientNet-B4 pada sistem deteksi nilai mata uang kertas rupiah pada berbagai kondisi visual.
3. Menganalisis pengaruh penggantian *backbone* YOLOv5 dari CSPDarknet ke EfficientNet-B4 terhadap kualitas deteksi dan lokalisasi objek pada berbagai kondisi visual.

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Memberikan pengetahuan mengenai dampak perubahan arsitektur *backbone* terhadap kinerja sistem deteksi objek.
2. Hasil penelitian dapat dimanfaatkan sebagai acuan teknis pada penelitian selanjutnya dalam pemilihan *backbone* pada pengembangan sistem deteksi objek, khususnya dalam menyeimbangkan kebutuhan akurasi dan keterbatasan sumber daya komputasi.

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Tinjauan pustaka berfungsi sebagai landasan teoritis sekaligus sarana untuk menegaskan kebaruan penelitian. Oleh karena itu, bagian ini mengkaji beberapa penelitian terdahulu yang relevan dan memiliki hubungan dengan topik penelitian yang dibahas.

Perkembangan deteksi objek modern semakin pesat setelah diterapkannya metode *deep learning*. Deteksi nilai mata uang merupakan salah satu bentuk penerapan deteksi objek yang tugasnya tidak sederhana. Sistem harus mampu menentukan lokasi objek secara tepat, menghasilkan *bounding box* yang presisi, serta tetap akurat meskipun citra diambil dalam berbagai kondisi, seperti pencahayaan berbeda, posisi miring, dan terlipat. Kesalahan kecil dalam lokalisasi dapat menyebabkan nilai mata uang terdeteksi keliru. Oleh karena itu, sistem deteksi uang membutuhkan model yang stabil, akurat, dan mampu menangkap detail fitur dengan baik.

Sejalan dengan kebutuhan tersebut, Jiao et al. [8] menerapkan Convolutional Neural Network (CNN) untuk mengenali uang kertas yang terlipat guna mengatasi keterbatasan perangkat transaksi otomatis. Penelitian tersebut menggunakan dataset citra uang kertas Ukraina yang terdiri dari uang normal, uang terlipat, dan uang rusak dengan total 24.000 citra. Model CNN 9-layer yang dikembangkan mampu mencapai akurasi tertinggi sebesar 96,46%, lebih baik

dibandingkan model LeNet-5 dengan akurasi 95,56% pada epoch 100. Selain itu, penelitian juga menunjukkan bahwa ukuran input citra dan konfigurasi parameter jaringan memengaruhi performa model secara signifikan. Ketahanan model terhadap kondisi uang yang terlipat menunjukkan bahwa CNN mampu mempertahankan performa pengenalan objek meskipun kondisi visual objek tidak ideal. Namun, penelitian tersebut masih berfokus pada klasifikasi nominal uang dan belum membahas lokalisasi objek menggunakan *bounding box* dalam konteks *object detection*. Evaluasi performa juga hanya menggunakan metrik akurasi tanpa mempertimbangkan metrik lokalisasi seperti mAP@0.5 dan mAP@0.5:0.95 yang lebih representatif untuk mengukur ketepatan deteksi objek. Selain itu, kebutuhan dataset yang besar dan kompleksitas jaringan masih menjadi tantangan dalam penerapan sistem secara *real-time*. Oleh karena itu, diperlukan model deteksi yang tidak hanya mampu mengenali fitur visual uang, tetapi juga memiliki kemampuan lokalisasi objek yang akurat dan efisien secara komputasi pada berbagai kondisi visual.

Rafi et al melalui penelitian berjudul *NSTU-BDTAKA: An Open Dataset for Bangladeshi Paper Currency Detection and Recognition* menunjukkan bahwa variasi kondisi fisik uang, perubahan pencahayaan, latar belakang yang kompleks, serta sudut pengambilan gambar berpengaruh langsung terhadap kinerja model deteksi [1]. Penelitian tersebut menggunakan YOLOv5 untuk mendeteksi dan mengenali uang kertas Bangladesh pada berbagai kondisi lingkungan nyata. Penelitian tersebut menyebutkan bahwa kualitas gambar yang tidak konsisten, oklusi objek, perubahan intensitas cahaya, kebutuhan pemrosesan *real-time*, variasi

ukuran uang, dan keterbatasan data pelatihan dapat menurunkan efektivitas dan presisi sistem deteksi, bahkan menyebabkan kesalahan klasifikasi maupun *false alarm*. Temuan tersebut menunjukkan bahwa sistem deteksi uang tidak hanya memerlukan kemampuan klasifikasi yang baik, tetapi juga membutuhkan kemampuan ekstraksi fitur dan lokalisasi objek yang lebih kuat agar tetap stabil pada berbagai kondisi visual.

Dibandingkan dengan penelitian Rafi et al, penelitian ini berfokus pada optimasi arsitektur model dengan mengganti *backbone* default YOLOv5 menggunakan EfficientNet-B4 untuk meningkatkan kualitas ekstraksi fitur dan presisi lokalisasi objek. Selain itu, penelitian ini mengevaluasi performa model secara lebih komprehensif menggunakan precision, recall, mAP@0.5, dan mAP@0.5:0.95 sehingga kontribusi penelitian tidak hanya pada implementasi deteksi uang, tetapi juga pada analisis peningkatan kualitas deteksi dan ketepatan *bounding box*.

Nasayreh et al. [9] dalam penelitiannya mengusulkan metode untuk mendeteksi uang kertas palsu Yordania menggunakan kombinasi *Convolutional Neural Network* (CNN) dan mekanisme *attention*. Mekanisme *attention* digunakan untuk memfokuskan model pada fitur-fitur penting dan mengurangi pengaruh fitur yang kurang relevan sehingga proses ekstraksi fitur menjadi lebih efektif. Penelitian ini menggunakan dataset uang Yordania dengan lima nominal serta menerapkan augmentasi data melalui penyesuaian kecerahan gambar untuk meningkatkan variasi data pelatihan. Hasil penelitian menunjukkan performa yang tinggi dengan accuracy sebesar 96%, precision 96,6%, recall 96,4%, dan F1-score 94,5%. Temuan

tersebut menunjukkan bahwa model berbasis *deep learning* dengan mekanisme *attention* mampu meningkatkan ketelitian dalam membedakan uang asli dan palsu melalui ekstraksi fitur visual yang lebih fokus dan representatif.

Berbeda dengan penelitian Nasayreh et al yang menggunakan mekanisme *attention* untuk memfokuskan model pada fitur-fitur penting dan relevan, pada penelitian ini menggunakan backbone EfficientNet-B4 pada YOLOv5 untuk meningkatkan kemampuan ekstraksi fitur sejak tahap awal jaringan. EfficientNet-B4 tidak hanya memfokuskan perhatian pada fitur tertentu, tetapi juga menghasilkan representasi fitur yang lebih kaya dan detail melalui peningkatan kedalaman, lebar, dan resolusi jaringan secara seimbang melalui *compound scaling*. Dengan demikian, EfficientNet-B4 diharapkan mampu mengekstraksi pola visual uang Rupiah secara lebih mendalam sehingga kualitas deteksi dan lokalisasi objek dapat meningkat. Selain itu penelitian Nasayreh et al masih berfokus pada proses klasifikasi citra dan belum membahas kemampuan lokalisasi objek menggunakan *bounding box* dalam konteks *object detection*. Pada penelitian ini menekankan kemampuan lokalisasi objek secara akurat melalui *bounding box*. Evaluasi performa tidak hanya menggunakan precision dan recall, tetapi juga mAP@0.5 serta mAP@0.5:0.95 yang lebih representatif dalam mengukur kualitas deteksi dan ketepatan lokalisasi objek. Dengan demikian, kontribusi penelitian ini tidak hanya pada pengenalan nominal uang, tetapi juga pada peningkatan kualitas deteksi dan presisi lokalisasi objek dalam sistem deteksi uang berbasis *real-time*.

Xu et al. dalam penelitiannya menjelaskan bahwa YOLOv5 merupakan algoritma *object detection* yang populer dan banyak digunakan di berbagai bidang,

termasuk industri dan kendaraan otonom. Namun, algoritma ini masih memiliki kelemahan berupa munculnya *false positive* dan *false negative* saat mendeteksi objek berukuran kecil [10]. Hal tersebut terjadi karena informasi detail pada objek kecil sering hilang selama proses ekstraksi fitur pada layer yang lebih dalam sehingga memengaruhi ketepatan klasifikasi maupun lokalisasi objek. Kondisi tersebut dapat menyebabkan *bounding box* yang dihasilkan menjadi kurang presisi, yang tercermin dari penurunan nilai *mean Average Precision* (mAP) sebagai indikator kualitas lokalisasi objek. Untuk mengatasi masalah tersebut, Xu et al. melakukan optimasi pada struktur YOLOv5s melalui modifikasi bagian *detection head* dan *loss function*. Penelitian tersebut menambahkan *multilevel feature fusion*, *decoupled attention detection head*, dan *focal MPDIoU loss* untuk meningkatkan kemampuan deteksi objek kecil serta mengurangi kesalahan *false positive* dan *false negative*. Hasil eksperimen menunjukkan peningkatan nilai mAP@0.5 pada beberapa dataset, seperti BDD100K yang meningkat dari 38,2% menjadi 44%, CCTSDB dari 81% menjadi 83,1%, dan GTSDB dari 95,8% menjadi 99%. Selain itu, nilai mAP@0.5:0.95 juga mengalami peningkatan, yang menunjukkan perbaikan kualitas lokalisasi objek dan ketepatan *bounding box*. Penelitian tersebut juga menggunakan metrik precision dan recall untuk mengevaluasi performa model secara lebih komprehensif. Model menunjukkan proses konvergensi yang lebih cepat, meskipun peningkatan performa tersebut diikuti dengan kenaikan jumlah parameter model dari 7,2 juta menjadi 14,6 juta serta penurunan kecepatan inferensi.

Berbeda dengan penelitian tersebut yang berfokus pada optimasi *detection head* dan *loss function*, penelitian ini melakukan optimasi pada bagian backbone YOLOv5 menggunakan EfficientNet-B4 untuk meningkatkan kemampuan ekstraksi fitur secara lebih mendalam. Jika penelitian Xu et al. meningkatkan sensitivitas deteksi objek kecil melalui penguatan aliran fitur dan mekanisme perhatian pada *detection head*, maka penelitian ini berfokus pada peningkatan kualitas representasi fitur sejak tahap awal ekstraksi fitur di backbone. EfficientNet-B4 dipilih karena menerapkan *compound scaling* yang menyeimbangkan kedalaman, lebar, dan resolusi jaringan sehingga mampu menghasilkan fitur visual yang lebih kaya dan representatif.

Perkembangan YOLO terus mengalami penyempurnaan untuk meningkatkan akurasi dan efisiensi komputasi, dan salah satu versi yang banyak digunakan adalah YOLOv5 karena memiliki keseimbangan yang baik antara kecepatan dan ketelitian deteksi. Namun, untuk meningkatkan performa lebih lanjut, beberapa peneliti mencoba menggabungkan keunggulan berbagai model. Seperti yang dilakukan oleh Meena et al [7] mengusulkan penerapan pendekatan *hybrid* melalui penggabungan empat model deteksi berbasis *Convolutional Neural Network* (CNN), yaitu *Single Shot Detector* (SSD), *You Only Look Once* (YOLO), RetinaNet, dan Faster R-CNN, yang mampu menghasilkan tingkat akurasi yang lebih tinggi dalam pengenalan objek multi-label dibandingkan dengan metode konvensional. Pendekatan *Hybrid Neural Network Architecture* merupakan strategi perancangan jaringan saraf yang mengintegrasikan berbagai model atau teknik *machine learning* untuk memanfaatkan keunggulan masing-masing arsitektur,

sehingga keterbatasan yang terdapat pada satu model tunggal dapat diminimalkan dan kinerja sistem secara keseluruhan dapat ditingkatkan.

Penelitian lain oleh Nurmaini dan Pratama [11] menganalisis pengaruh penggunaan backbone CSPResNeXt-50, CSPDarknet53, dan EfficientNet-B0 terhadap kinerja YOLOv4. Hasil penelitian menunjukkan bahwa CSPDarknet53, dengan ukuran model paling besar (244,2 MB), memperoleh nilai mAP keseluruhan tertinggi sebesar 83,41% sehingga menghasilkan performa deteksi yang lebih stabil dibandingkan CSPResNeXt-50 dan EfficientNet-B0. Namun, pada threshold IoU yang sangat ketat yaitu 0.95, EfficientNet-B0 memperoleh nilai Average Precision (AP) terbaik sebesar 4,84%, lebih tinggi dibandingkan CSPDarkNet53 sebesar 4,26% dan CSPResNeXt-50 sebesar 0,91%. Hasil ini menunjukkan bahwa meskipun CSPDarkNet53 unggul pada performa rata-rata keseluruhan, EfficientNet-B0 memiliki kemampuan lokalisasi objek yang lebih presisi pada kondisi evaluasi yang ketat. Pada IoU 0.95, prediksi *bounding box* prediksi hampir sama dengan *ground truth*, sehingga nilai AP pada *threshold* ini mencerminkan kualitas lokalisasi objek secara lebih detail. Selain itu, penelitian tersebut juga menunjukkan bahwa *backbone* dengan struktur yang lebih kompleks cenderung meningkatkan kualitas ekstraksi fitur, namun dapat berdampak pada efisiensi komputasi dan waktu inferensi. Temuan ini menjadi salah satu dasar dalam penelitian ini untuk memilih EfficientNet sebagai pengganti *backbone* default YOLOv5 yaitu CSPDarkNet53. Dengan menggunakan EfficientNet-B4 yang memiliki kapasitas ekstraksi fitur lebih besar dibanding EfficientNet-B0, penelitian ini diharapkan mampu meningkatkan kualitas representasi fitur dan presisi

lokalisasi objek uang rupiah, terutama pada evaluasi mAP@0.5:0.95. Penelitian ini juga memperkuat bahwa komponen *backbone* memiliki peran penting dalam menentukan kualitas deteksi, ketepatan *bounding box*, dan kinerja keseluruhan sistem YOLO.

Penelitian Nugroho dan Baihaqi [3] mengusulkan penggantian backbone standar YOLOv5 dengan MobileNetV3 Small untuk mendeteksi atribut sekolah secara *real-time*. Hasil eksperimen menunjukkan bahwa model yang telah dimodifikasi memperoleh nilai mAP@0.5 sebesar 91.2%, lebih tinggi dibandingkan YOLOv5 yang memperoleh nilai 88.2%. Selain itu, precision meningkat dari 89.2% menjadi 91.4% dan recall meningkat dari 81.9% menjadi 87.3%. Peningkatan nilai mAP@0.5 tersebut menunjukkan bahwa penggantian backbone mampu meningkatkan kemampuan model dalam mengenali objek sekaligus memperbaiki kualitas lokalisasi objek dan presisi *bounding box*. Temuan tersebut relevan dengan penelitian ini karena sama-sama melakukan optimasi YOLOv5 melalui penggantian backbone. Namun, berbeda dengan penelitian Nugroho dan Baihaqi yang menggunakan MobileNetV3 Small untuk meningkatkan efisiensi model, penelitian ini menggunakan EfficientNet-B4 untuk meningkatkan kemampuan ekstraksi fitur dan kualitas representasi visual objek uang rupiah. Selain itu, penelitian ini tidak hanya menggunakan mAP@0.5, tetapi juga mAP@0.5:0.95 untuk mengevaluasi kualitas lokalisasi objek secara lebih ketat.

Penelitian Mahasin et al [12] membandingkan penggunaan backbone CSPDarkNet53, CSPResNeXt-50, dan EfficientNet-B0 pada YOLOv4 untuk mendeteksi mikrofosil foraminifera. Hasil eksperimen menunjukkan bahwa

CSPDarkNet53 memperoleh nilai mAP tertinggi sebesar 83,41% dengan FPS 32, sehingga menghasilkan performa deteksi paling baik dari sisi akurasi keseluruhan dan kecepatan inferensi. Nilai mAP yang tinggi menunjukkan bahwa model mampu melakukan klasifikasi dan lokalisasi objek dengan baik melalui ketepatan *bounding box*. Sementara itu, CSPResNeXt-50 memperoleh precision, recall, dan F1-score tertinggi yaitu masing-masing sebesar 75,60%, 81,10%, dan 78%, sedangkan EfficientNet-B0 memperoleh nilai mAP sebesar 81,76% dengan ukuran model paling ringan yaitu 156,8 MB. Menariknya, meskipun nilai mAP EfficientNet-B0 lebih rendah dibanding CSPDarkNet53, *backbone* ini memperoleh nilai Average Precision (AP) terbaik pada IoU 0.95 sebesar 4,84%, lebih tinggi dibanding CSPDarkNet53 sebesar 4,26% dan CSPResNeXt-50 sebesar 0,91%. Hasil ini menunjukkan bahwa EfficientNet-B0 memiliki kemampuan lokalisasi objek yang lebih presisi pada threshold IoU yang sangat ketat, dimana *bounding box* prediksi hampir sama dengan *ground truth*. Temuan tersebut mengindikasikan bahwa backbone EfficientNet mampu menghasilkan representasi fitur yang lebih detail untuk menentukan posisi objek secara lebih akurat.

Penelitian ini relevan dengan penelitian yang dilakukan karena sama-sama menganalisis pengaruh backbone terhadap performa sistem deteksi objek. Namun, berbeda dengan penelitian Mahasin et al yang menggunakan YOLOv4 dan EfficientNet-B0 pada dataset mikrofosil, sedangkan penelitian ini menggunakan YOLOv5 dengan backbone EfficientNet-B4 pada dataset uang rupiah. Selain mengevaluasi precision, recall, dan mAP@0.5, penelitian ini juga menggunakan mAP@0.5:0.95 untuk menilai kualitas lokalisasi objek secara lebih ketat.

Penelitian Li et al [13] menggunakan kombinasi EfficientNet dan YOLOv5 untuk mendeteksi kerusakan pipa bawah laut pada citra dengan kondisi visual yang kompleks seperti blur, noise, dan kontras rendah. Hasil penelitian menunjukkan bahwa EfficientNet memperoleh akurasi klasifikasi sebesar 93,57%, sedangkan YOLOv5 mencapai nilai mAP sebesar 92.42% dan FPS 25. Penelitian tersebut menunjukkan bahwa EfficientNet memiliki kemampuan ekstraksi fitur yang baik pada kondisi citra tidak ideal. Hal ini relevan dengan penelitian yang dilakukan karena deteksi uang rupiah juga menghadapi tantangan berupa variasi pencahayaan, lipatan uang, dan latar belakang kompleks. Namun, berbeda dengan penelitian Li et al yang menggunakan EfficientNet untuk klasifikasi dan YOLOv5 untuk deteksi, penelitian ini secara khusus mengevaluasi EfficientNet-B4 sebagai backbone pengganti CSPDarkNet pada YOLOv5 untuk meningkatkan kualitas lokalisasi objek melalui evaluasi mAP@0.5 dan mAP@0.5:0.95.

Beberapa penelitian menunjukkan bahwa pemilihan *backbone* sangat memengaruhi kualitas fitur yang dihasilkan serta ketepatan model dalam menentukan posisi objek. *Backbone* dengan struktur yang lebih dalam dan kaya fitur berpotensi meningkatkan kemampuan model dalam menghasilkan *bounding box* yang lebih presisi, bahkan ketika citra diambil dalam berbagai kondisi. Namun, penggunaan backbone yang lebih kompleks juga berdampak pada meningkatnya kebutuhan komputasi, waktu pelatihan, dan waktu inferensi model.

Dobrzycki et al. [14] mengkaji permasalahan tersebut melalui analisis strategi *layer-freezing* pada arsitektur YOLO dalam skema *transfer learning*. Hasil penelitian menunjukkan bahwa pengaturan layer yang dibekukan (*freeze*) dan layer

yang dilatih ulang (*fine-tuning*) berpengaruh terhadap stabilitas pelatihan, kecepatan konvergensi, efisiensi penggunaan GPU, serta performa akhir model dalam deteksi objek. Strategi *freezing* yang tepat memungkinkan model mempertahankan fitur umum hasil pretraining pada layer awal backbone, seperti tekstur, tepi, dan pola visual dasar, sekaligus menyesuaikan layer yang lebih dalam terhadap karakteristik dataset baru secara lebih efisien.

Temuan tersebut mendukung penelitian ini yang menerapkan transfer learning pada YOLOv5 dengan mengganti backbone CSPDarknet menjadi EfficientNet-B4 serta menggunakan strategi freeze–unfreeze training. Pendekatan ini diharapkan mampu mempertahankan kemampuan ekstraksi fitur umum dari model pretrained sekaligus meningkatkan adaptasi terhadap karakteristik visual uang rupiah yang memiliki detail, tekstur, warna, dan ornamen yang kompleks. Dengan demikian, performa deteksi tidak hanya dipengaruhi oleh pemilihan backbone, tetapi juga oleh strategi transfer learning yang digunakan selama proses pelatihan model.

Berdasarkan penelusuran penelitian terdahulu, belum ditemukan penelitian yang menggunakan dataset yang identik dengan dataset pada penelitian ini. Oleh karena itu, perbandingan dilakukan terhadap penelitian sejenis yang membahas deteksi nilai mata uang maupun peningkatan performa object detection melalui modifikasi backbone, optimasi arsitektur YOLO, dan penerapan transfer learning. Perlu diperhatikan bahwa setiap penelitian menggunakan dataset yang berbeda sehingga nilai performa yang diperoleh tidak dapat dibandingkan secara langsung. Oleh karena itu, selain menampilkan nilai performa, Tabel 2.1 juga menyajikan

informasi mengenai dataset yang digunakan pada masing-masing penelitian sebagai konteks dalam interpretasi hasil penelitian terdahulu.

Tabel 2.1 Ringkasan Penelitian Terdahulu

No	JuduL Penelitian	Algoritma	Dataset	Karakteristik Visual	Performance
1.	Comparison of CSPDarkNet53, CSPResNeXt-50, and EfficientNet-B0 Backbones on YOLO V4 as Object Detector. Marsa Mahasin, Irma Amelia Dewi, IJESTY, 2022	YOLOv4 dengan backbone: ▪ CSPDarkNet53 ▪ CSPResNeXt-50 EfisienNet-B0	Mikrofosil batuan	Ukuran dataset sangat kecil	▪ CSPDarkNet53 mAP@0.5:0.95: 83.41% ▪ CSPResNeXt-50 mAP@0.5:0.95: 81.0% ▪ EfisienNet-B0 mAP@0.5:0.95: 81.76%
2.	Deep Learning-Based Detection of Fake Multinational Banknotes in a Cross-Dataset Environment Utilizing Smartphone Cameras for Assisting Visually Impaired Individuals Pham et al, MDPI, 2022.	YOLOv4	Uang kertas USD, EUR, KRW dan JOD	Variasi pencahayaan, sudut, uang terlipat, kusut dan background kompleks	▪ KRW mAP@0.5 : 56.713% ▪ USD mAP@0.5 : 60.253% ▪ JOD mAP@0.5 : 40.769% ▪ EUR mAP@0.5 : 78.8%
3.	Implementation of Deep Learning using Convolutional Neural Network Method and Yolo Algorithm in Rupiah Banknote Detection System. Mulyani et al, https://doi.org/10.23887/jet.v9i3.98189, 2025	YOLOv4	Uang kertas rupiah	Kusut dan terlipat	mAP@0.5 : 88.0%

4..	Improved Small Object Detection Algorithm CRL-YOLOv5. Wang et al, Sensors 2024, 24, 6437.https://doi.org/10.3390/s24196437h, 2024	YOLOv5	Dataset VisDrone2019	dataset objek kecil	mAP@0.5 : 39,2% mAP@0.5:0.95: 22.3%
5.	A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS Terven et al, MDPI, 2023	Studi review yang menganalisis perkembangan YOLOv1 hingga YOLOv8 dan YOLO-NAS			YOLOv5: mAP@0.5:0.95: 55.8
6.	Object detection using YOLO: challenges, architectural successors, datasets and applications. Terven et al, MDPI, 2023	Analisis perbandingan di antara berbagai versi YOLO	MS COCO		YOLOv4: mAP@0.5:0.95: 65.7
7.	A Decade of You Only Look Once (YOLO) for Object Detection: A Review. Ramos and Sappa, IEEE Access, 2025	Review YOLOv1 hingga YOLOv13	MS COCO		YOLOv5: mAP@0.5 : 56.8% mAP@0.5:0.95: 37.4%
8.	Performance Evaluation of YOLOv5-based Custom Object Detection Model for Campus-Specific Scenario Dontabhaktuni Jayakumar and Samineni Peddakrishna, https://doi.org/10.52756/ijerr.2024.v38.005, 2024	YOLOv5 dengan variasi Batch size dan Epoch	Eksperimen dilakukan pada dataset yang beragam yang terdiri dari berbagai objek: pejalan kaki, kendaraan, bangunan, dan rintangan.		Batch Size: 32 Epoch: 100 mAP@0.5 : 82.1% mAP@0.5:0.95: 50.4%

Meskipun berbagai penelitian telah mengkaji peningkatan performa YOLO melalui modifikasi arsitektur maupun penggunaan backbone alternatif, masih terdapat kesenjangan penelitian yang perlu dikaji lebih lanjut. Berdasarkan studi literatur yang telah dilakukan, belum ditemukan penelitian yang secara khusus mengevaluasi penggantian backbone CSPDarknet pada YOLOv5 menggunakan EfficientNet-B4 untuk deteksi nilai mata uang rupiah dengan fokus pada peningkatan kualitas ekstraksi fitur dan ketepatan lokalisasi objek. Selain itu, penelitian terdahulu umumnya berfokus pada peningkatan akurasi deteksi secara umum dan belum banyak menganalisis kualitas lokalisasi objek menggunakan metrik yang lebih ketat, seperti mAP@0.5:0.95. Di sisi lain, penerapan strategi pelatihan freeze–unfreeze dalam skema penggantian backbone juga belum banyak dibahas secara sistematis untuk mengevaluasi pengaruhnya terhadap proses transfer learning, stabilitas pelatihan, performa deteksi, serta efisiensi komputasi model. Oleh karena itu, penelitian ini mengusulkan penggunaan EfficientNet-B4 sebagai backbone alternatif pada YOLOv5 yang dikombinasikan dengan strategi pelatihan freeze–unfreeze untuk meningkatkan kualitas ekstraksi fitur dan presisi lokalisasi objek pada sistem deteksi nilai mata uang rupiah.

2.2 Keaslian Penelitian

Tabel 2.2 Matriks literatur review dan posisi penelitian
Analisis Eksperimental Penggantian *Backbone* YOLOv5 Dengan Efficientnet-B4 Pada Sistem Deteksi Nilai Mata Uang

No	Judul. Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
1.	NSTU-BDTAKA: An open dataset for Bangladeshi paper currency detection and recognition	Md. Jubayar Alam Rafi, Mohammad Rony, Nazia Majadi, Elsevier, 2024	Penelitian ini menggunakan metode pengembangan dataset dengan pengumpulan citra uang kertas Bangladesh secara langsung di berbagai kondisi lingkungan. Dataset disusun untuk dua tugas, yaitu deteksi menggunakan YOLOv5 dan pengenalan nominal, dengan anotasi *bounding box* serta penerapan augmentasi data untuk meningkatkan keragaman citra.	Hasil penelitian berupa dataset terbuka NSTU-BDTAKA yang terdiri dari ribuan citra teranotasi untuk deteksi dan pengenalan uang kertas. Evaluasi awal menggunakan YOLOv5 menunjukkan bahwa dataset mampu mendukung pengujian model pada kondisi visual yang beragam dan mendekati skenario penggunaan nyata.	Keunggulan penelitian ini adalah tersedianya dataset uang kertas dengan variasi kondisi fisik dan pencahayaan yang realistis, sehingga relevan untuk penelitian deteksi dan pengenalan uang kertas. Kelemahannya adalah keterbatasan dataset meliputi variasi kualitas citra, perubahan pencahayaan, adanya occlusion, perbedaan ukuran uang kertas, keterbatasan data latih, serta belum adanya perbandingan lintas mata uang dan pengujian terhadap gangguan visual.	Penelitian ini berfokus pada penyediaan dataset dan evaluasi awal model tanpa modifikasi arsitektur, sedangkan tesis ini menganalisis penggantian *backbone* YOLOv5 dengan EfficientNet-B4 dan dampaknya terhadap akurasi deteksi, ketelitian lokalisasi *bounding box*, serta efisiensi komputasi pada sistem deteksi nilai mata uang rupiah.

2.	Rupiah Banknotes Detection: Comparison of The Faster R-CNN Algorithm and YOLOv5	Zuhdi Hanif et al, Jurnal Infotel Vol. 16, No. 3, August 2024, PP. 502–517, 2024	Penelitian ini menggunakan metode eksperimen komparatif dengan membandingkan dua algoritma deteksi objek, yaitu Faster R-CNN dan YOLOv5, untuk mendeteksi uang kertas rupiah. Evaluasi dilakukan menggunakan metrik akurasi deteksi dan waktu pemrosesan untuk menilai perbedaan performa kedua algoritma.	Hasil penelitian menunjukkan bahwa YOLOv5 memiliki keunggulan dari sisi kecepatan pemrosesan dan efisiensi komputasi dibandingkan Faster R-CNN, sementara akurasi deteksi kedua metode relatif sebanding untuk skenario deteksi uang kertas rupiah.	Keunggulan penelitian ini adalah analisis perbandingan dua algoritma deteksi populer pada kasus nyata uang kertas rupiah. Kelemahannya adalah penelitian ini belum menganalisis pengaruh variasi kondisi lingkungan dan fisik uang kertas secara mendalam serta belum mengevaluasi ketelitian lokalisasi dan stabilitas performa model pada berbagai skenario.	Penelitian ini berfokus pada perbandingan algoritma deteksi objek (Faster R-CNN dan YOLOv5) tanpa memodifikasi struktur internal YOLOv5. Sebaliknya, tesis ini menitikberatkan pada analisis eksperimental penggantian *backbone* YOLOv5 dengan EfficientNet-B4 dan mengevaluasi dampaknya terhadap akurasi deteksi nominal rupiah, ketelitian lokalisasi *bounding box*, serta efisiensi komputasi.
3.	Implementasi Deep Learning Menggunakan Metode *Convolutional Neural Network* Dan Algoritma Yolo Dalam Sistem Pendeteksi Uang Kertas Rupiah Bagi Penyandang Low Vision	Maulana Azhar dkk., TRANSIENT, VOL. 10, NO. 3. 2021	Penelitian ini menerapkan metode *Convolutional Neural Network* (CNN) dan algoritma YOLO untuk membangun sistem pendeteksi uang kertas rupiah yang ditujukan bagi penyandang low vision. Sistem dirancang untuk mengenali nominal uang kertas melalui	Hasil penelitian menunjukkan bahwa sistem mampu mendeteksi dan mengenali nominal uang kertas rupiah serta membantu penyandang low vision dalam mengenali uang secara mandiri. YOLO digunakan sebagai metode deteksi utama dengan performa	Keunggulan penelitian ini adalah penerapan sistem deteksi uang kertas yang berorientasi pada kebutuhan pengguna penyandang low vision dan integrasi dengan keluaran audio. Kelemahannya adalah pengujian masih dilakukan pada kondisi terbatas, belum membahas variasi kondisi lingkungan secara luas seperti pencahayaan ekstrem dan	Penelitian ini berfokus pada implementasi sistem aplikasi berbasis YOLO untuk membantu penyandang low vision tanpa melakukan analisis mendalam terhadap arsitektur model. Sebaliknya, tesis ini menitikberatkan pada analisis eksperimental penggantian *backbone* YOLOv5 dengan EfficientNet-B4 dan mengevaluasi dampaknya

			citra dan memberikan output berupa informasi suara sebagai bantuan pengguna.	yang cukup baik pada kondisi pengujian terbatas.	occlusion, serta belum menganalisis ketelitian lokalisasi *bounding box* dan stabilitas performa model secara mendalam.	terhadap akurasi deteksi nilai mata uang rupiah, ketelitian lokalisasi *bounding box*, serta efisiensi komputasi.
4.	Jordanian banknote data recognition: A CNN-based approach with attention mechanism	Nasayreh et al, Elsevier, 2024	Penelitian ini mengembangkan metode pengenalan uang kertas Yordania dengan memanfaatkan kombinasi *Convolutional Neural Network* (CNN) dan mekanisme attention untuk mengidentifikasi serta membedakan uang kertas asli dan palsu secara lebih akurat melalui penekanan pada fitur-fitur visual yang relevan.	Hasil penelitian menunjukkan bahwa integrasi mekanisme attention pada CNN mampu meningkatkan akurasi pengenalan dan pembedaan uang kertas asli dan palsu dibandingkan pendekatan CNN konvensional, karena model lebih fokus pada fitur visual penting yang bersifat diskriminatif.	Keunggulan penelitian ini terletak pada penggunaan mekanisme attention yang meningkatkan kemampuan model dalam mengekstraksi fitur penting uang kertas. Kelemahannya adalah penelitian berfokus pada tugas pengenalan/klasifikasi dan tidak membahas proses deteksi objek, ketelitian lokalisasi *bounding box*, maupun evaluasi kinerja sistem secara real-time.	Penelitian ini berfokus pada pengenalan dan klasifikasi keaslian uang kertas menggunakan CNN dengan attention mechanism. Sebaliknya, tesis ini menitikberatkan pada sistem deteksi objek menggunakan YOLOv5 dengan *backbone* EfficientNet-B4 untuk mendeteksi dan melokalisasi nilai mata uang rupiah, serta menganalisis akurasi deteksi, ketelitian *bounding box*, dan efisiensi komputasi.
5.	YOLOv3 Based Currency Detection and Recognition System for Visually Impaired Person	Joshi et al, IEEE., 2020	Membuat sistem yang dapat mendeteksi dan mengenali uang kertas secara real-time untuk membantu penyandang	Hasil penelitian menunjukkan bahwa sistem berbasis YOLOv3 mampu mendeteksi dan mengenali nominal uang kertas dengan akurasi deteksi	Keunggulan penelitian ini adalah penerapan sistem deteksi uang kertas yang berorientasi pada kebutuhan penyandang tunanetra dan pemanfaatan YOLOv3 yang mendukung	Penelitian ini menggunakan YOLOv3 sebagai algoritma deteksi untuk membangun sistem bantu pengenalan uang kertas. Sebaliknya, tesis ini menggunakan YOLOv5 dengan *backbone*

			tunanetra dalam bertransaksi. Sistem ini dirancang untuk bekerja secara mandiri tanpa memerlukan koneksi internet.	sebesar 95,71%. Sistem ini bekerja secara mandiri dan berfungsi secara real-time, sehingga efektif digunakan oleh individu dengan gangguan penglihatan.	pemrosesan relatif cepat. Kelemahannya adalah evaluasi sistem yang masih terbatas pada kondisi lingkungan tertentu, belum dianalisis secara mendalam pengaruh variasi pencahayaan, occlusion, sudut pengambilan gambar, serta kondisi fisik uang kertas, dan belum disajikan analisis kuantitatif mengenai ketelitian lokalisasi *bounding box* maupun performa real-time secara menyeluruh.	EfficientNet-B4 dan berfokus pada analisis eksperimental pengaruh penggantian *backbone* terhadap akurasi deteksi nilai mata uang rupiah, ketelitian lokalisasi *bounding box*, dan efisiensi komputasi.
6.	Folding Paper Currency Recognition and Research Based on Convolution Neural Network	Mengshu Jiao, IEEE, 2018	Penelitian ini mengembangkan metode pengenalan uang kertas terlipat menggunakan *Convolutional Neural Network* (CNN). Model dilatih untuk mengenali nominal uang kertas dengan mengekstraksi fitur visual dari citra uang yang mengalami lipatan,	Hasil penelitian menunjukkan bahwa pendekatan CNN mampu mengenali uang kertas dalam kondisi terlipat dengan tingkat akurasi yang cukup baik, serta menunjukkan potensi CNN dalam menangani distorsi bentuk akibat lipatan pada uang kertas.	Keunggulan penelitian ini adalah fokus khusus pada permasalahan pengenalan uang kertas terlipat yang jarang dibahas dalam penelitian lain. Kelemahannya adalah ruang lingkup penelitian yang terbatas pada kondisi lipatan tertentu, pendekatan yang hanya mencakup tugas klasifikasi tanpa deteksi objek dan lokalisasi *bounding box*, serta belum	Penelitian ini berfokus pada pengenalan nominal uang kertas terlipat berbasis klasifikasi CNN. Sebaliknya, tesis ini menitikberatkan pada sistem deteksi objek menggunakan YOLOv5 dengan *backbone* EfficientNet-B4 untuk mendeteksi dan melokalisasi nilai mata uang rupiah dalam berbagai kondisi fisik, serta menganalisis kinerja

			sehingga sistem tetap mampu melakukan pengenalan meskipun bentuk uang tidak utuh.		adanya evaluasi performa sistem pada kondisi lingkungan yang lebih kompleks atau secara real-time.	deteksi, ketelitian *bounding box*, dan efisiensi komputasi.
7.	Currency Recognition For The Visually Impaired People	Yihao Wang, IEEE, 2023	Penelitian ini bertujuan mengembangkan model pengenalan mata uang yang mampu mengidentifikasi dan mengklasifikasikan denominasi uang kertas dari berbagai negara melalui aplikasi mobile berbasis Android. Sistem pengenalan mata uang otomatis dirancang khusus untuk membantu individu dengan gangguan penglihatan dengan memanfaatkan teknik pengolahan citra dan algoritma machine learning guna mengenali	Hasil penelitian menunjukkan bahwa sistem pengenalan mata uang berbasis visi komputer yang diimplementasikan pada perangkat mobile mampu mengklasifikasikan berbagai denominasi mata uang dari beberapa negara dengan tingkat akurasi yang cukup baik, serta memberikan informasi yang membantu penyandang tunanetra dalam mengenali uang kertas secara mandiri.	Keunggulan penelitian ini adalah pengembangan aplikasi mobile yang mendukung pengenalan mata uang lintas negara dan berorientasi pada kebutuhan pengguna penyandang tunanetra. Kelemahannya adalah penelitian lebih berfokus pada tugas klasifikasi denominasi tanpa analisis mendalam terhadap proses deteksi objek dan lokalisasi *bounding box*, serta pengujian sistem yang masih terbatas pada kondisi lingkungan tertentu.	Penelitian ini menekankan pengenalan dan klasifikasi denominasi mata uang melalui aplikasi mobile. Sebaliknya, tesis ini menitikberatkan pada sistem deteksi objek menggunakan YOLOv5 dengan *backbone* EfficientNet-B4 untuk mendeteksi dan melokalisasi nilai mata uang rupiah, serta menganalisis akurasi deteksi, ketelitian *bounding box*, dan efisiensi komputasi.

			denominasi uang kertas secara akurat.			
8.	Classification of Rupiah to Help Blind with The *Convolutional Neural Network* Method	Ery Pamungkas dkk, Shinta2, 2022	Penelitian ini menerapkan metode *Convolutional Neural Network* (CNN) untuk mengklasifikasikan nominal uang kertas rupiah sebagai alat bantu bagi penyandang tunanetra. Sistem bekerja dengan mengenali citra uang kertas dan menentukan kelas nominal berdasarkan fitur visual yang dipelajari oleh CNN.	Hasil penelitian menunjukkan bahwa model CNN mampu mengklasifikasikan nominal uang kertas rupiah dengan tingkat akurasi yang cukup baik pada data uji yang digunakan, sehingga sistem dapat membantu penyandang tunanetra dalam mengenali uang kertas secara mandiri.	Keunggulan penelitian ini adalah penerapan metode CNN yang sederhana dan efektif untuk klasifikasi nominal uang kertas rupiah serta orientasi sistem pada kebutuhan penyandang tunanetra. Kelemahannya adalah pendekatan yang hanya mencakup tugas klasifikasi tanpa proses deteksi objek dan lokalisasi *bounding box*, pengujian yang masih terbatas pada kondisi tertentu, serta belum dievaluasi ketahanan sistem terhadap variasi pencahayaan dan kondisi fisik uang kertas yang beragam.	Penelitian ini berfokus pada klasifikasi nominal uang kertas rupiah berbasis CNN. Sebaliknya, tesis ini menitikberatkan pada sistem deteksi objek menggunakan YOLOv5 dengan *backbone* EfficientNet-B4 untuk mendeteksi dan melokalisasi nilai mata uang rupiah, serta menganalisis akurasi deteksi, ketelitian *bounding box*, dan efisiensi komputasi.
9.	Indian Currency Detection using Image Recognition Technique	Kalpna Gautam, IEEE *Xplore*, 2020	Penelitian ini mengembangkan metode pengenalan citra untuk mendeteksi dan mengenali uang kertas India serta membedakan antara uang asli dan palsu	Hasil penelitian menunjukkan bahwa teknik pengenalan citra yang digunakan mampu mendeteksi dan mengenali nominal uang kertas India serta membedakan uang	Keunggulan penelitian ini adalah kemampuannya dalam mendeteksi dan mengenali uang kertas sekaligus membedakan keaslian uang menggunakan pendekatan pengenalan citra yang relatif sederhana.	Penelitian ini berfokus pada pengenalan dan verifikasi keaslian uang kertas berbasis teknik pengenalan citra. Sebaliknya, tesis ini menitikberatkan pada sistem deteksi objek menggunakan YOLOv5

			dengan mengekstraksi fitur visual khas dari citra uang kertas dan melakukan klasifikasi berdasarkan karakteristik tersebut.	asli dan palsu dengan tingkat akurasi yang cukup baik pada kondisi pengujian tertentu.	Kelemahannya adalah metode yang digunakan belum memanfaatkan arsitektur deep learning modern secara mendalam, pengujian masih terbatas pada kondisi lingkungan tertentu, serta belum dievaluasi secara komprehensif pada skenario real-time dan variasi kondisi fisik uang kertas.	dengan *backbone* EfficientNet-B4 untuk mendeteksi dan melokalisasi nilai mata uang rupiah, serta menganalisis akurasi deteksi, ketelitian *bounding box*, dan efisiensi komputasi.
10.	Improved Small Object Detection Algorithm Based on YOLOv5	Bo Xu, Bin Gao, Yunhu Li, IEEE Computer Society, 2024	Penelitian ini mengusulkan peningkatan performa YOLOv5 untuk mendeteksi objek berukuran kecil dengan melakukan modifikasi pada struktur jaringan dan strategi ekstraksi fitur, termasuk peningkatan fusi fitur multi-skala dan penyesuaian komponen deteksi agar lebih sensitif terhadap objek kecil.	Model YOLOv5 yang ditingkatkan menunjukkan peningkatan akurasi deteksi objek kecil dibandingkan YOLOv5 standar, dengan penurunan kecepatan yang relatif kecil.	Keunggulan penelitian ini adalah peningkatan performa deteksi objek kecil melalui modifikasi arsitektur internal. Kelemahannya adalah meningkatnya kompleksitas model dan potensi beban komputasi tambahan.	Penelitian ini menunjukkan peningkatan performa deteksi objek kecil terutama pada metrik akurasi deteksi, namun dengan konsekuensi kompleksitas model yang lebih tinggi. Sedangkan tesis ini mengevaluasi peningkatan mAP YOLOv5 dengan *backbone* EfficientNet-B4, yang sama-sama berdampak pada peningkatan kompleksitas model.

11.	An Analysis of Layer-Freezing Strategies for Enhanced Transfer Learning in YOLO Architectures	Dobrzycki et al, MDPI, 2025	Penelitian ini menggunakan eksperimen kuantitatif dengan menguji berbagai strategi layer-freezing pada model YOLO pretrained dan mengevaluasinya menggunakan mAP, kurva loss, serta penggunaan memori GPU.	Hasil penelitian menunjukkan bahwa strategi layer-freezing yang tepat meningkatkan stabilitas pelatihan dan efisiensi komputasi, di mana system freezing menghasilkan performa deteksi sebanding dengan fine-tuning penuh namun dengan kebutuhan sumber daya lebih rendah.	Keunggulan penelitian ini terletak pada analisis strategi pelatihan dan transfer learning secara sistematis.	Penelitian ini menitikberatkan pada strategi layer-freezing pada YOLO, sedangkan tesis ini menganalisis penggantian *backbone* YOLOv5 dengan EfficientNet-B4 menggunakan skema freezing–unfreezing serta dampaknya terhadap akurasi deteksi, ketelitian *bounding box*, dan efisiensi komputasi.
12.	Money Recognition for the Visually Impaired: A Case Study on Sri Lankan Banknotes	Akshaan Bandara, arXiv, 2025	Penelitian ini menggunakan EfficientDet-lite2 dengan *transfer learning* dan *data augmentation* untuk mendeteksi uang kertas Sri Lanka secara *real-time* berbasis *object detection*.	Model memperoleh AP 98.47%, AP50 99.83%, dan mAP 98.77% serta mampu bekerja secara *real-time* pada perangkat mobile.	Keunggulan penelitian ini adalah model ringan dan efisien untuk perangkat mobile serta mampu mendeteksi uang pada berbagai kondisi pencahayaan dan latar belakang. Kelemahannya adalah belum menganalisis pengaruh backbone terhadap kualitas lokalisasi objek dan efisiensi komputasi secara mendalam.	Penelitian ini sama-sama membahas deteksi uang kertas menggunakan *object detection* dan *transfer learning*. Namun, penelitian tersebut menggunakan EfficientDet-lite2, sedangkan tesis ini menggunakan YOLOv5 dengan backbone EfficientNet-B4 serta menganalisis pengaruh backbone dan strategi *freeze–unfreeze* terhadap mAP@0.5.

						mAP@0.5:0.95, dan efisiensi komputasi.
13.	Deep Learning Approach Combining Lightweight CNN Architecture with Transfer Learning: An Automatic Approach for the Detection and Recognition of Bangladeshi Banknotes	Ali Hasan Md. Linkon et al, IEEE, 2020	Penelitian ini menggunakan pendekatan *transfer learning* dengan arsitektur CNN ringan seperti MobileNet, NASNetMobile, dan ResNet152v2 untuk mendeteksi dan mengenali uang kertas Bangladesh.	Hasil penelitian menunjukkan bahwa MobileNet memperoleh akurasi 98,88% pada dataset 8000 citra, NASNetMobile mencapai 100% pada dataset 1970 citra, dan MobileNet memperoleh 97,77% pada dataset gabungan.	Keunggulan penelitian ini adalah penggunaan model ringan berbasis *transfer learning* yang efisien dan cocok untuk perangkat mobile. Kelemahannya adalah penelitian masih berfokus pada klasifikasi/pengenalan uang kertas dan belum membahas deteksi objek serta ketelitian lokalisasi *bounding box*.	Penelitian ini sama-sama menggunakan *transfer learning* dan dataset uang kertas. Namun, penelitian tersebut berfokus pada klasifikasi menggunakan CNN ringan, sedangkan tesis ini menggunakan YOLOv5 dengan backbone EfficientNet-B4 untuk mendeteksi dan melokalisasi nilai mata uang rupiah serta menganalisis mAP@0.5, mAP@0.5:0.95, dan efisiensi komputasi model.
14.	BankNote-Net: Open Dataset for Assistive Universal Currency Recognition	Felipe Oviedo et al. arXiv, 20221	Penelitian ini mengembangkan dataset BankNote-Net dan menggunakan MobileNetV2 dengan supervised contrastive learning untuk pengenalan uang kertas dari 17 mata uang dan 112 denominasi.	Dataset terdiri dari 24.826 citra uang kertas dengan berbagai kondisi pencahayaan, orientasi, blur, dan latar belakang. Model menunjukkan performa yang baik pada skenario transfer learning dan few-shot learning.	Keunggulan penelitian ini adalah penggunaan dataset besar dan beragam serta penerapan transfer learning pada berbagai mata uang. Kelemahannya adalah penelitian masih berfokus pada pengenalan/klasifikasi uang dan belum membahas ketelitian lokalisasi objek menggunakan bounding box dan mAP.	Penelitian ini sama-sama menggunakan transfer learning dan backbone ringan berbasis CNN. Namun, penelitian tersebut berfokus pada klasifikasi universal berbagai mata uang, sedangkan penelitian ini menggunakan YOLOv5 dengan backbone EfficientNet-B4 untuk mendeteksi dan melokalisasi uang rupiah serta menganalisis

						mAP@0.5, mAP@0.5:0.95, dan efisiensi komputasi model.
15.	Adaptive RoI-aware Network for Accurate Banknote Recognition Using Natural Images	Wang et al, Soft Computing, 2025	Penelitian ini mengusulkan Adaptive RoI-aware Network dengan multiscale deformable convolution dan non-local attention untuk mengenali uang kertas pada berbagai kondisi citra alami seperti perubahan skala, sudut, pencahayaan, dan latar belakang kompleks.	Model memperoleh akurasi sebesar 92,29% pada dataset INR, 97,42% pada PLN, 99,63% pada BTN, dan 99,64% pada UAE serta menunjukkan kemampuan generalisasi yang baik pada berbagai mata uang.	Keunggulan penelitian ini adalah kemampuan model dalam memfokuskan area penting uang seperti watermark, serial number, dan pola khusus sehingga meningkatkan akurasi pada kondisi visual yang kompleks. Kelemahannya adalah penelitian masih berfokus pada pengenalan/klasifikasi uang dan belum membahas deteksi objek menggunakan bounding box dan evaluasi mAP.	Penelitian ini sama-sama membahas ekstraksi fitur pada citra uang kertas dengan kondisi visual yang kompleks. Namun, penelitian tersebut menggunakan pendekatan RoI-aware dan attention untuk klasifikasi uang, sedangkan penelitian ini menggunakan YOLOv5 dengan backbone EfficientNet-B4 untuk mendeteksi dan melokalisasi uang rupiah serta menganalisis mAP@0.5, mAP@0.5:0.95, dan efisiensi komputasi model.

2.3 Landasan Teori

2.3.1 Uang Kertas Rupiah

Uang kertas rupiah adalah alat pembayaran yang sah di Indonesia yang diterbitkan oleh Bank Indonesia (BI). Uang kertas rupiah terdiri atas beberapa nilai nominal rupiah yang dilengkapi dengan detail dan fitur spesifik yang berbeda untuk masing-masing nilai mata uang. Berikut adalah salah satu contoh detail pecahan uang kertas rupiah tahun emisi 2022 disertai detail objeknya:

a. Rp. 1.000: Pahlawan Tjut Meutia, Tari Tifa, Banda Neira dan Bunga Anggrek Larat
b. Rp. 2.000: Pahlawan M. Husni Thamrin, Tari Piring, Ngarai Sianok, dan Bunga Jeumpa
c. Rp. 5.000: Pahlawan DR. K. H. Idham Chalid, Tari Gambyong, Gunung Bromo dan Bunga Sedap Malam
d. Rp. 10.000: Pahlawan Frans Kaisiepo, Tari Pakarena, Taman Nasional Wakatobi, dan Bunga Cempaka Hutan Kasar
e. Rp. 20.000: Pahlawan DR. G.S.S.J Ratulangi, Tari Gong, Derawan, dan Bunga Anggrek Hitam.
f. Rp. 50.000: Pahlawan Ir. H. Djuanda Kartawidjaja, Tari Legong, Taman Nasional Komodo, dan Bunga Jepun Bali.
g. Rp. 100.000: Pahlawan Ir. Soekarno dan Drs. Mohammad Hatta, Topeng Betawi, Raja Ampat dan Bunga Anggrek.

Selain warna, detail serta fitur spesifik pada setiap pecahan uang kertas rupiah dimanfaatkan dalam sistem deteksi objek untuk meningkatkan akurasi dalam mengenali dan mengklasifikasikan masing-masing pecahan secara tepat.

Gambar 2.1 Gambar uang kertas rupiah

Hingga saat ini, penelitian tentang pengenalan mata uang terus berkembang pesat, dengan berbagai algoritma dan pendekatan yang diuji untuk mencapai performa prediksi gambar yang lebih tinggi. Para peneliti dari berbagai negara telah mengeksplorasi beragam teknik machine learning dan deep learning, mulai dari metode klasik seperti Support Vector Machines (SVM) hingga arsitektur neural network yang lebih canggih seperti *Convolutional Neural Networks* (CNN) dan variasinya. Setiap pendekatan memiliki kelebihan dan tantangannya masing-masing, dengan fokus pada peningkatan kinerja model, kecepatan pemrosesan, dan kemampuan generalisasi terhadap berbagai kondisi pengambilan gambar. Perkembangan teknik tersebut kemudian mendorong munculnya model-model Hybrid Neural Network Architecture yang memadukan model YOLOv5 dengan arsitektur modern lain seperti EfficientNet yang diharapkan dapat meningkatkan performa deteksi objek.

2.3.2 *Convolutional Neural Network*

Convolutional Neural Network (CNN) merupakan arsitektur *deep learning* yang dirancang khusus untuk pengolahan data citra dengan memanfaatkan operasi konvolusi sebagai mekanisme utama ekstraksi fitur. Dalam dunia kecerdasan buatan, khususnya bidang pengolahan citra (*image processing*), *Convolutional Neural Network* (CNN) menjadi salah satu arsitektur jaringan saraf tiruan yang paling populer. CNN dirancang khusus untuk mengenali pola dalam gambar seperti tepi, tekstur, bentuk, dan objek dengan efisiensi yang tinggi. Kemampuan ini membuat CNN sangat efektif dalam berbagai tugas seperti klasifikasi gambar, pengenalan objek, dan segmentasi citra.

Arsitektur CNN bekerja dengan cara mengekstrak fitur-fitur penting dari citra masukan melalui beberapa jenis layer, seperti *convolution layer*, *pooling layer*, dan *fully connected layer*. *Convolution layer* berfungsi untuk mengekstraksi fitur dasar dari gambar, misalnya tepi, tekstur, atau pola tertentu. *Pooling layer* digunakan untuk mengurangi ukuran *feature map* sehingga proses komputasi menjadi lebih efisien. Setelah itu, hasil ekstraksi fitur akan diproses oleh *fully connected layer* untuk melakukan proses klasifikasi objek.

Proses ini membantu jaringan memahami elemen-elemen dalam citra, seperti pola sederhana di lapisan awal dan pola kompleks di lapisan yang lebih dalam. Pada Gambar 2.2 di bawah ini, ditampilkan struktur arsitektur CNN klasik, dimana citra masukan diolah melalui beberapa tahapan hingga menghasilkan prediksi kelas citra.

Gambar 2.2 Arsitektur CNN Klasik

Pada arsitektur tersebut, citra masukan pertama kali diproses oleh *convolution layer* untuk mengekstraksi ciri-ciri dasar dari citra. Lapisan ini bekerja dengan menerapkan filter berukuran kecil yang digeser ke seluruh area citra, sehingga mampu menangkap informasi lokal citra. Hasil dari proses ini berupa *feature map* yang merepresentasikan respons jaringan terhadap pola tertentu dalam citra.

Selanjutnya, *feature map* tersebut dilewatkan ke *pooling layer* yang berfungsi untuk mengurangi ukuran data tanpa menghilangkan informasi penting. Proses ini membantu menurunkan kompleksitas komputasi serta meningkatkan ketahanan model terhadap variasi kecil pada posisi objek dalam citra. Tahapan konvolusi dan *pooling* dapat diulang beberapa kali untuk memperkaya representasi fitur yang dipelajari jaringan. Setelah seluruh fitur diekstraksi, hasilnya diratakan dan diteruskan ke *fully connected layer*. Lapisan ini menggabungkan fitur-fitur yang diperoleh untuk menentukan kelas citra. Pada tahap akhir, jaringan menghasilkan prediksi kelas berdasarkan pola yang dipelajari selama proses pelatihan.

Secara umum, proses pada CNN dapat dinyatakan sebagai fungsi transformasi dari data input menjadi representasi fitur sebagai berikut:

$$Y_i = F_i(X_i) \quad (1)$$

Dimana:

X_i : citra input

F_i : fungsi transformasi jaringan CNN

Y_i : hasil ekstraksi fitur atau prediksi model

Dalam CNN model tersusun dari beberapa lapisan yang saling terhubung dan bekerja secara berurutan untuk mengekstraksi fitur dari citra. Setiap lapisan pada jaringan memproses informasi dari lapisan sebelumnya sehingga jaringan mampu mempelajari berbagai pola visual secara bertahap, mulai dari pola sederhana hingga pola yang lebih kompleks.

Citra input pada CNN biasanya direpresentasikan dalam bentuk tensor dengan dimensi (H_i, W_i, C_i) di mana H_i dan W_i menunjukkan ukuran citra (tinggi dan lebar), sedangkan C_i menunjukkan jumlah kanal warna atau jumlah fitur pada citra. Representasi tensor ini memungkinkan jaringan saraf memproses informasi visual dalam bentuk matriks multi dimensi.

Dalam arsitektur CNN, model dapat dipandang sebagai rangkaian fungsi berlapis yang saling terhubung [12]. Pada setiap tahap jaringan, suatu lapisan dapat diulang beberapa kali untuk meningkatkan kemampuan ekstraksi fitur. Secara matematis, struktur jaringan dapat dinyatakan sebagai berikut:

$$N = \odot F_i^L(X_{(H_i, W_i, C_i)})_{i=1\ldots n} \quad (2)$$

Dimana:

F_i^L: menunjukkan bahwa fungsi pada tahap ke-i diulang sebanyak L kali

$\odot$: menunjukkan menunjukkan komposisi operasi fungsi pada setiap tahap jaringan

$X_{(H_i, W_i, C_i)}$: menunjukkan tensor input dengan dimensi tinggi, lebar, dan jumlah kanal.

Melalui proses ini, jaringan CNN dapat mengekstraksi berbagai fitur penting dari citra secara bertahap, mulai dari pola sederhana hingga pola yang lebih kompleks.

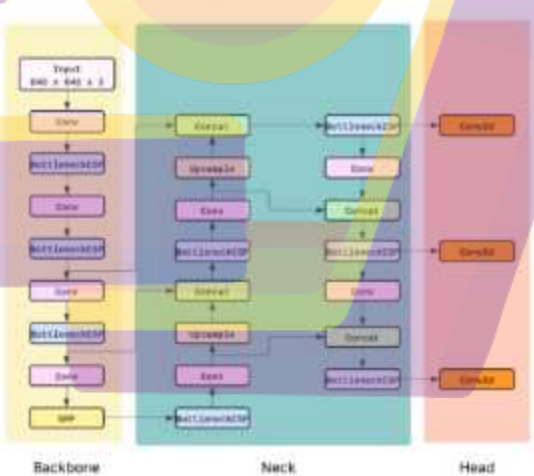
Kinerja dan kompleksitas CNN umumnya dipengaruhi oleh tiga faktor utama yaitu:

- *Depth* (kedalaman jaringan) menunjukkan jumlah lapisan dalam jaringan yang menentukan seberapa dalam model dapat mengekstraksi fitur dari gambar. Semakin banyak lapisan yang digunakan, semakin kompleks pola yang dapat dipelajari oleh jaringan.
- *Width* (lebar Jaringan) menunjukkan jumlah kanal fitur pada setiap lapisan yang mempengaruhi kapasitas model dalam menangkap variasi fitur pada citra. Semakin banyak kanal yang digunakan, semakin banyak informasi visual yang dapat diproses dalam satu lapisan.
- *Resolution* (resolusi input) menunjukkan ukuran citra yang diproses oleh model. Resolusi yang lebih tinggi memungkinkan jaringan menangkap detail visual yang lebih banyak sehingga membantu model dalam mengenali objek dengan lebih baik.

Melalui proses ini, CNN mampu mengenali pola visual yang kompleks dan sering digunakan pada berbagai aplikasi pengolahan citra seperti pengenalan wajah, klasifikasi objek, dan sistem deteksi objek. Berdasarkan prinsip kerja tersebut, CNN menjadi fondasi utama bagi model deteksi objek modern, termasuk YOLOv5, yang mengintegrasikan proses ekstraksi fitur dan prediksi objek dalam satu arsitektur terpadu.

2.3.3 Model YOLOv5

YOLOv5 atau "You Only Look Once" versi ke-5 merupakan generasi lanjutan dari model YOLO yang dirancang untuk memberikan performa lebih baik dalam mendeteksi objek secara *real-time* dengan akurasi tinggi dan kecepatan pemrosesan yang efisien. Arsitektur YOLOv5 dikembangkan dengan desain yang fleksibel dan ringan, sehingga dapat digunakan pada perangkat keras dengan sumber daya terbatas seperti smartphone dan IoT devices.



Gambar 2.3 Arsitektur YOLOv5

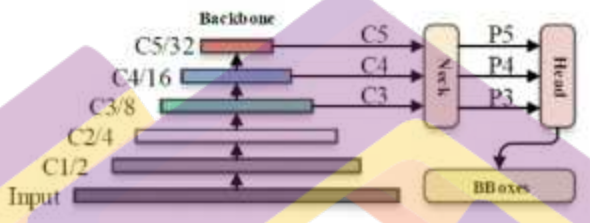
Arsitektur YOLOv5 seperti ditunjukkan Gambar 2.3 terdiri dari tiga komponen utama, yaitu *Backbone*, Neck, dan Head, yang masing-masing memiliki peran penting dalam mendeteksi objek pada citra [15].

Pada YOLOv5, *backbone* yang digunakan adalah CSPDarknet53. *Backbone* berperan sebagai bagian dari *Convolutional Neural Network* yang bertugas mengekstraksi fitur-fitur penting dalam gambar dan diolah secara bertahap. CSPDarknet53 berperan dalam memisahkan dan menggabungkan informasi gradien secara iteratif, serta mengintegrasikan perubahan gradien ke dalam *feature map*. Hal ini berdampak pada peningkatan akurasi dan efisiensi model serta mengurangi ukuran model dengan mengurangi jumlah parameter [16]. Pada YOLOv5, citra input terlebih dahulu diproses oleh bagian *backbone* untuk mengekstraksi fitur penting dari gambar. Backbone menghasilkan *feature map* yang merepresentasikan karakteristik visual objek.

Feature map tersebut kemudian diproses oleh bagian neck yang terletak di tengah-tengah antara *backbone* dan head. Neck bertindak sebagai penghubung antara keduanya [17]. Neck bertugas menggabungkan informasi fitur dari berbagai skala menggunakan konsep Feature Pyramid Network (FPN) untuk memastikan deteksi yang akurat.

Tahap terakhir adalah *detection* head, yang bertugas memprediksi lokasi objek serta kelas objek yang terdeteksi. Head terdiri dari tiga cabang, masing-masing berfungsi untuk memprediksi fitur pada skala yang berbeda. Setiap cabang menghasilkan *boanding box, class probability,* dan *confidence score*. Proses

mengubah gambar input menjadi *boanding box* disebut proses inferensi. Gambar 2.3 berikut menggambarkan proses inferensi default dari YOLOv5 [16].



Gambar 2.4 Diagram Alir Inferensi Default YOLOv5

Dalam hal deteksi nilai mata uang, melalui Gambar 2.4 dijelaskan bagaimana prosesnya dimulai dari Input berupa gambar uang kertas rupiah yang dimasukkan ke dalam sistem model YOLOv5 dalam resolusi 640×640 piksel. Sistem akan membaca elemen-elemen penting seperti angka nominal, warna, dan fitur keamanan yang terdapat pada uang kertas. Pada tahap awal, gambar diproses oleh *backbone* yang berfungsi untuk mengekstraksi fitur utama dari citra menggunakan arsitektur CSPDarknet53.

Proses ekstraksi fitur terdiri dari beberapa blok konvolusi Conv (*Convolution Layer*) dan modul *Bottleneck* CSP yang berfungsi meningkatkan efisiensi komputasi sekaligus mempertahankan kualitas representasi fitur. Dalam backbone ini terdapat beberapa tahap ekstraksi fitur yang direpresentasikan oleh layer C1, C2, C3, C4, dan C5 sebagai berikut:

- Layer C1/C2 menangkap informasi dasar seperti warna, tekstur, dan pola sederhana dari uang kertas.
- Layer C3/C4 pola yang lebih kompleks seperti angka nominal, logo Bank Indonesia, serta beberapa elemen keamanan pada uang kertas.

- Layer C5 mengekstraksi fitur tingkat tinggi yang merepresentasikan kombinasi pola unik pada setiap denominasi uang kertas.

Setelah fitur penting diekstraksi oleh *backbone*, fitur-fitur ini diteruskan ke bagian Neck yang menggunakan *Feature Pyramid Network* (FPN) dan *Path Aggregation Network* (PANet) untuk menggabungkan fitur dari berbagai level ekstraksi yaitu level tinggi (C5), menengah (C4), dan rendah (C3). Pada tahap ini dilakukan operasi *Upsample* untuk meningkatkan resolusi *Feature map* serta *Concat* untuk menggabungkan *Feature map* dari level yang berbeda. Proses ini juga diperkuat oleh blok Conv dan BottleneckCSP untuk memperkaya representasi fitur. Melalui proses ini, sistem dapat menggabungkan informasi fitur tingkat rendah, menengah, dan tinggi sehingga model mampu mendeteksi objek dengan ukuran yang berbeda.

Hasil dari proses ini adalah *Feature map* pada tiga skala berbeda yang dikenal sebagai P3, P4, dan P5, yang masing-masing merepresentasikan objek kecil, sedang, dan besar. Dengan demikian, model dapat mengenali uang kertas dalam berbagai skala, orientasi, dan posisi, serta membantu sistem membedakan nominal uang meskipun memiliki warna atau pola yang mirip.

Pada bagian Head, *Feature map* P3, P4, dan P5 digunakan untuk menghasilkan prediksi akhir melalui layer Conv2D sebagai detection layer. Pada tahap ini model menghasilkan tiga jenis keluaran utama, yaitu:

- Lokasi *Bounding* yang menunjukkan posisi objek dalam gambar.

$$B = (x,y,w,h) \quad (3)$$

Dimana:

x,y : Koordinat pusat objek relatif terhadap grid.

w,h: Lebar dan tinggi kotak pembatas relatif terhadap ukuran gambar

- *Confidence score* untuk menunjukkan tingkat keyakinan model terhadap keberadaan objek tersebut. Nilai *Confidence score* berkisar antara 0%-100%. Jika *Confidence score* rendah, makak *bounding box* tersebut kemungkinan salah deteksi (*false positive*) dan bisa dihapus.

$$\text{Confidence} = P(\text{Object}) \cdot IOU_{pred}^{truth} \quad (4)$$

Dimana:

P (Object): Probabilitas ada objek dalam grid.

IOU_{pred}^{truth} : Intersection Over Union antara *bounding box* prediksi dan ground truth (ukuran seberapa akurat prediksi).

- *Class Probability* untuk mengidentifikasi denominasi uang dengan menghitung probabilitas untuk setiap kategori yang telah dilatih. Kelas dengan probabilitas tertinggi akan dipilih sebagai hasil akhir. Di sini, model terkadang dapat menghasilkan prediksi yang keliru apabila perbedaan nilai probabilitas antar kelas sangat tipis atau ketika kualitas fitur yang diekstraksi kurang jelas, sehingga sistem memilih kelas yang tidak sesuai dengan objek sebenarnya. Kondisi ini semakin diperparah oleh keterbatasan dalam presisi dan lokalisasi, terutama pada objek berukuran kecil, karena pembagian grid yang tetap terkadang terlalu besar sehingga tidak mampu menangkap detail fitur objek secara optimal dan menyebabkan *bounding box* yang dihasilkan kurang presisi. Selain itu, ketika dua objek berada dalam satu grid yang sama atau saling

tumpang tindih, model dapat mengalami kesulitan dalam membedakan serta menentukan posisi masing-masing objek secara akurat.

YOLOv5 menggunakan *Non-Maximum Suppression* (NMS) untuk menyaring prediksi yang tidak relevan atau saling tumpang tindih dengan cara memilih *bounding box* yang memiliki *confidence score* tertinggi dan menghapus kotak lain yang memiliki nilai IoU (*Intersection over Union*) tinggi terhadap kotak tersebut. IoU dihitung sebagai perbandingan antara area irisan dan area gabungan dua *bounding box*.

$$IOU = \frac{area\ irisan}{area\ gabungan}$$

Meskipun metode ini efektif untuk mengurangi duplikasi deteksi, NMS memiliki keterbatasan, terutama ketika dua objek berada sangat berdekatan atau saling tumpang tindih, karena sistem dapat menghapus *bounding box* yang sebenarnya valid. Selain itu, jika *bounding box* awal sudah kurang presisi atau salah posisi, NMS tidak dapat memperbaiki kesalahan lokalisasi tersebut, melainkan hanya memilih kotak terbaik dari prediksi yang tersedia.

Pada tahap akhir deteksi, sistem menghasilkan keluaran berupa daftar objek yang teridentifikasi, yang mencakup lokasi *bounding box*, nilai *confidence score*, dan kelas objek atau denominasi uang yang diprediksi. Lokasi *bounding box* menunjukkan posisi objek dalam citra, *confidence score* menggambarkan tingkat keyakinan model terhadap keberadaan objek tersebut, sedangkan kelas objek menunjukkan hasil klasifikasi berdasarkan probabilitas tertinggi. Namun, apabila tahapan sebelumnya, seperti proses ekstraksi fitur atau deteksi pada kondisi

pencahayaan yang kurang baik, tidak berjalan secara optimal, maka hasil akhir yang diperoleh dapat menjadi kurang akurat baik dari sisi klasifikasi maupun presisi lokalisasinya.

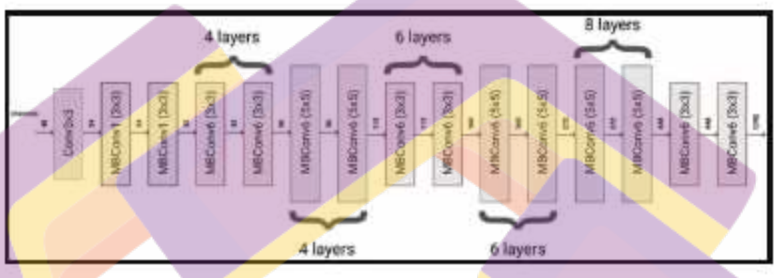
2.3.4 EfficientNet-B4

EfficientNet merupakan arsitektur *Convolutional Neural Network* yang dikembangkan untuk meningkatkan kinerja model dengan cara memperbesar jaringan secara efisien. Pada banyak penelitian sebelumnya, peningkatan performa CNN biasanya dilakukan dengan memperbesar satu dimensi jaringan saja, misalnya menambah jumlah layer atau memperbesar resolusi gambar. Pendekatan tersebut sering kali menghasilkan model yang lebih kompleks tetapi tidak selalu memberikan peningkatan akurasi yang optimal.

EfficientNet memperkenalkan pendekatan yang disebut *compound scaling*, yaitu metode penskalaan jaringan yang meningkatkan tiga komponen utama secara bersamaan, yaitu kedalaman jaringan, lebar jaringan, dan resolusi citra input. Dengan meningkatkan ketiga komponen tersebut secara seimbang, model dapat meningkatkan kemampuan ekstraksi fitur tanpa menyebabkan peningkatan komputasi yang berlebihan.

EfficientNet pada awalnya dikembangkan untuk tugas klasifikasi citra. Arsitektur ini berfungsi mengekstraksi fitur dari gambar dan menghasilkan prediksi kelas objek tanpa menentukan lokasi objek dalam citra. Oleh karena itu, EfficientNet tidak memiliki mekanisme untuk menghasilkan *bounding box* ataupun mendeteksi beberapa objek dalam satu gambar secara langsung. Dalam aplikasi deteksi objek, EfficientNet umumnya digunakan sebagai *backbone* atau ekstraktor

fitur yang kemudian dikombinasikan dengan algoritma deteksi objek, seperti YOLO, yang memiliki kemampuan untuk menentukan lokasi objek dan melakukan klasifikasi secara bersamaan.



Gambar 2.4 Arsitektur Jaringan EfficientNet

Arsitektur jaringan EfficientNet-B4 seperti ditunjukkan pada Gambar 2.4 [18] merupakan salah satu arsitektur CNN yang dirancang untuk meningkatkan kinerja model dengan menyeimbangkan tiga faktor utama, yaitu kedalaman jaringan, lebar jaringan, dan resolusi input. EfficientNet-B4 tersusun atas beberapa blok MBConv (*Mobile Inverted Bottleneck Convolution*) yang berfungsi mengekstraksi fitur penting dari citra secara bertahap, mulai dari pola visual sederhana hingga pola yang lebih kompleks.

Tan dan Le mengembangkan arsitektur EfficientNet, salah satunya EfficientNet-B4 yang digunakan dalam penelitian ini. Tan dan Le mengusulkan metode penskalaan jaringan yang disebut *compound scaling* untuk meningkatkan akurasi sekaligus menjaga efisiensi komputasi model CNN [19]. Metode ini meningkatkan ukuran jaringan secara seimbang dengan menggunakan koefisien penskalaan ϕ (phi) yang mengatur peningkatan pada *depth*, *width*, dan *resolution* secara bersamaan.

Compound scaling mempengaruhi kualitas *Feature map* melalui tiga aspek utama antara lain:

1. Peningkatan Depth (Kedalaman Jaringan)

   Depth yang lebih besar berarti jaringan memiliki lebih banyak layer konvolusi. Secara sederhana dapat dituliskan:

   $$d = \alpha^{\phi} \quad (5)$$

   Semakin besar nilai φ, maka semakin dalam jaringan CNN yang digunakan dalam arsitektur EfficientNet. Kedalaman jaringan yang lebih besar berarti jumlah layer konvolusi pada jaringan juga meningkat. Dampaknya pada *Feature map* yaitu layer yang lebih banyak memungkinkan jaringan untuk mengekstraksi fitur secara lebih bertahap dan mendalam, sehingga model dapat menangkap pola visual yang lebih kompleks serta mempelajari struktur objek secara lebih detail. Proses ini meningkatkan kemampuan representasi fitur pada jaringan, sehingga *Feature map* yang dihasilkan menjadi lebih kaya informasi dan mampu menggambarkan karakteristik objek dengan lebih jelas untuk membantu proses deteksi objek secara lebih akurat.

2. Peningkatan *Width* (Jumlah *Channel Feature Map*)

   Width menunjukkan jumlah filter konvolusi pada setiap layer. Secara matematis:

   $$w = \beta^{\phi} \quad (6)$$

   Jika jumlah filter pada suatu layer meningkat, maka jumlah *channel* pada *Feature map* yang dihasilkan juga akan bertambah. Dampaknya pada *Feature map* yaitu peningkatan jumlah *channel* memungkinkan jaringan menyimpan

lebih banyak informasi visual dari citra yang diproses. Dengan jumlah *channel* yang lebih banyak, jaringan dapat mengekstraksi lebih banyak jenis fitur visual, mempelajari variasi pola yang lebih beragam, serta membedakan objek yang memiliki karakteristik yang mirip. Akibatnya, *Feature map* yang dihasilkan menjadi lebih representatif dalam menggambarkan ciri-ciri objek pada citra, sehingga membantu model dalam meningkatkan ketepatan proses deteksi dan klasifikasi objek.

3. Peningkatan Resolution (Resolusi Input)

   Resolution menunjukkan ukuran citra yang digunakan sebagai input jaringan. Secara matematis:

$$r = \gamma^{\phi} \tag{7}$$

   Resolusi yang lebih tinggi berarti citra memiliki jumlah piksel yang lebih banyak sehingga informasi visual pada gambar menjadi lebih lengkap. Dampaknya pada *Feature map* yaitu peningkatan resolusi memungkinkan jaringan mempertahankan detail visual yang lebih halus selama proses ekstraksi fitur. Dengan resolusi citra yang lebih tinggi, jaringan dapat mempertahankan detail objek yang lebih kecil, meningkatkan ketelitian dalam menentukan lokasi objek, serta membantu menghasilkan *bounding box* yang lebih akurat. Hal ini membuat *Feature map* yang dihasilkan mampu merepresentasikan objek dengan lebih jelas sehingga proses deteksi objek oleh model menjadi lebih tepat.

Dalam metode *compound scaling*, peningkatan ketiga dimensi tersebut tidak dilakukan secara bebas, tetapi mengikuti batasan tertentu agar keseimbangan antara

kedalaman jaringan, jumlah channel, dan resolusi input tetap terjaga. Batasan tersebut dinyatakan sebagai berikut:

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2, \quad \alpha \geq 1, \quad \beta \geq 1, \quad \gamma \geq 1 \tag{8}$$

Dimana:

Φ : koefisien penskalaan model yang menentukan seberapa besar kapasitas jaringan diperbesar

α : konstanta *scaling* yang mengatur peningkatan kedalaman jaringan

β : konstanta *scaling* yang mengatur peningkatan jumlah channel atau lebar jaringan

γ : konstanta *scaling* yang mengatur peningkatan resolusi citra input.

Batasan ini menunjukkan bahwa peningkatan kapasitas jaringan harus dilakukan secara seimbang antara depth, width, dan resolution agar kompleksitas komputasi jaringan tidak meningkat secara berlebihan namun tetap mampu meningkatkan akurasi model.

Persamaan-persamaan tersebut di atas menunjukkan bahwa peningkatan kapasitas jaringan dilakukan secara seimbang pada ketiga dimensi jaringan, sehingga model dapat menghasilkan *Feature map* yang lebih informatif dan detail. *Feature map* merupakan representasi fitur visual yang dihasilkan oleh backbone CNN setelah citra diproses melalui beberapa lapisan konvolusi. *Feature map* inilah yang kemudian digunakan oleh bagian deteksi YOLO untuk menentukan lokasi objek dan kelas objek.

Penerapan metode compound scaling pada EfficientNet memberikan dampak langsung terhadap peningkatan kinerja model. Hal ini terjadi karena tiga

dimensi utama jaringan, yaitu depth (kedalaman jaringan), width (jumlah channel), dan resolution (ukuran citra input) ditingkatkan secara bersamaan dan seimbang. Peningkatan ketiga aspek tersebut menyebabkan kualitas *Feature map* yang dihasilkan oleh jaringan menjadi lebih baik. *Feature map* yang lebih informatif memungkinkan jaringan menghasilkan representasi fitur objek yang lebih jelas, sehingga model menjadi lebih mudah dalam mengenali dan membedakan karakteristik objek pada citra. Secara umum, peningkatan kapasitas jaringan pada CNN juga berpengaruh terhadap kompleksitas komputasi model. Hubungan tersebut dapat dinyatakan sebagai berikut:

$$\text{FLOPS} \propto d \cdot w^2 \cdot r^2 \tag{9}$$

Persamaan tersebut menunjukkan bahwa kompleksitas komputasi jaringan (FLOPS) berbanding lurus dengan kedalaman jaringan (d), jumlah channel (w), dan resolusi citra input (r). Artinya, peningkatan kedalaman jaringan, jumlah channel, maupun resolusi citra akan menyebabkan kebutuhan komputasi model juga meningkat.

2.3.5 Metrik Evaluasi

Penelitian ini menerapkan metrik evaluasi yaitu *confusion matrix* yang lazim digunakan untuk menilai performa model, sehingga kinerja sistem deteksi objek dapat dianalisis secara objektif dan komprehensif. Ada 4 komponen yang digunakan yaitu (TP) *True Positive*, yaitu data yang positif dan diprediksi dengan benar, (FP) *False Positif*, yaitu data yang bernilai negatif tetapi diprediksi positif, (FN) *False Negatif*, yaitu data yang bernilai positif tetapi diprediksi sebagai negatif,

dan (TN) *True Negative* yaitu data yang negatif dan diprediksi dengan benar. Adapun metrik evaluasi yang digunakan antara lain:

a. Precision

Menggambarkan tingkat ketepatan model dalam memberikan prediksi positif. Metrik ini menunjukkan proporsi prediksi objek yang benar dibandingkan dengan seluruh prediksi objek yang dihasilkan oleh model. Dalam konteks penelitian ini, precision menunjukkan seberapa besar kemungkinan bahwa *bounding box* yang diprediksi sebagai uang kertas dengan nominal tertentu benar-benar merupakan uang kertas dengan kelas tersebut. Nilai precision yang tinggi menandakan bahwa model jarang menghasilkan deteksi keliru (*false positive*, sehingga hasil deteksi lebih dapat dipercaya.

$$Precision = \frac{TP}{(TP+FP)} \quad (10)$$

b. Recall

Mengukur kemampuan model dalam menemukan seluruh objek yang sebenarnya ada di dalam gambar. Pada sistem deteksi uang kertas, recall menunjukkan sejauh mana model mampu mendeteksi seluruh uang yang terdapat dalam gambar, termasuk pada kondisi sulit seperti pencahayaan redup, uang terlipat, dan kusam, atau latar belakang yang kompleks. Nilai recall yang rendah mengindikasikan masih adanya objek uang yang terlewatkan oleh sistem (*false negative*).

$$Recall = \frac{TP}{(TP+FN)} \quad (11)$$

c. Mean Average Precision (mAP)

mAP mengukur seberapa akurat model dalam menemukan dan mengenali objek yang benar, serta seberapa konsisten model tersebut dalam memberikan hasil yang tepat untuk setiap kelas. Dalam penelitian ini, evaluasi performa model tidak cukup hanya menggunakan metrik akurasi, karena akurasi hanya menunjukkan jumlah prediksi benar tanpa mempertimbangkan ketepatan posisi objek yang dideteksi. Pada system *object detection*, model harus mampu melakukan klasifikasi sekaligus menentukan lokasi objek melalui *bounding box*. Oleh karena itu, kualitas lokalisasi objek menjadi aspek penting dalam evaluasi performa model. Kualitas lokalisasi objek umumnya diukur menggunakan *Intersection over Union* (IoU), yaitu ukuran kesesuaian antara bounding box prediksi dengan ground truth. Apostolidis et al menjelaskan bahwa IoU digunakan untuk mengevaluasi *localization error* pada *object detection* dan memberikan gambaran mengenai ketepatan posisi *bounding box* hasil prediksi model [20]. Berdasarkan kondisi tersebut, penelitian ini menggunakan precision, recall, mAP@0.5, dan mAP@0.5:0.95 karena metrik tersebut lebih relevan dalam mengevaluasi kemampuan klasifikasi sekaligus kualitas lokalisasi objek pada sistem deteksi nilai mata uang rupiah. Fungsi keempat metrik tersebut:

- Precision: digunakan untuk mengukur ketepatan model dalam mendeteksi objek uang sehingga dapat menunjukkan jumlah *false positive* yang dihasilkan model.

- Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh objek uang pada citra sehingga dapat menunjukkan jumlah *false negative* yang berhasil dikurangi.
- mAP@0.5: mengukur akurasi deteksi objek dengan toleransi lokasi yang relatif longgar. Suatu prediksi dianggap benar apabila nilai *Intersection over Union (IoU)* antara *bounding box* prediksi dan *ground truth* minimal sebesar 0.5.

$$mAP@0.5 = \frac{1}{N} \sum_{i=1}^{N} AP_i^{IoU=0.5} \quad (12)$$

dimana N = jumlah kelas (nominal uang kertas)

- mAP@0.5:0.95 merupakan metrik evaluasi yang lebih ketat karena menghitung rata-rata Average Precision pada berbagai ambang IoU, mulai dari 0.5 hingga 0.95 dengan interval 0.05.

$$mAP@0.5{:}0.95 = \frac{1}{10N} \sum_{t=0.5}^{0.95} \sum_{i=1}^{N} AP_i^{IoU=t} \quad (13)$$

d. Inference time

Digunakan mengukur waktu yang dibutuhkan model untuk memproses satu gambar dan menghasilkan hasil deteksi. Metrik ini mencerminkan efisiensi komputasi dan menentukan kelayakan model untuk digunakan pada aplikasi *real-time*. Dalam penelitian ini, inference time digunakan untuk membandingkan kecepatan pemrosesan antara YOLOv5 baseline dan YOLOv5 dengan *backbone* EfficientNet-B4.

$$Inference\ Time = \frac{Total\ waktu\ inferensi}{jumlah\ citra\ yang\ diproses} \quad (14)$$

e. Improvement

Dalam penelitian ini dilakukan perbandingan antara model deteksi objek, YOLOv5 Baseline dan YOLOv5 dengan *backbone* EfficientNet-B4 untuk mengetahui seberapa besar peningkatan atau penurunan performa model setelah dilakukan modifikasi arsitektur. Selain itu, pengujian juga melibatkan *backbone* EfficientNet-B0 dan EfficientNet-B1 sebagai pembanding tambahan guna memberikan gambaran yang lebih komprehensif mengenai pengaruh variasi *backbone* EfficientNet terhadap kinerja deteksi objek. Berikut ukuran persentase peningkatan (*percentage improvement*) terhadap metrik evaluasi Precision, Recall, mAP@0.5, dan mAP@0.5:0.95.

$$\%\ Improvement = \frac{Nilai\ model\ baru - Nilai\ baseline}{Nilai\ baseline}\ x\ 100\% \qquad (15)$$

BAB 3
METODE PENELITIAN

3.1 Jenis, Sifat, dan Pendekatan Penelitian

Jenis dan pendekatan penelitian yang digunakan dalam penelitian ini adalah penelitian kuantitatif, karena analisis dilakukan berdasarkan perhitungan matematis terhadap data penelitian yang dihasilkan. Data tersebut diperoleh dari hasil pelatihan dan pengujian model deteksi nilai mata uang, yang selanjutnya dianalisis menggunakan metrik evaluasi kinerja secara numerik.

Berdasarkan sifatnya, penelitian ini termasuk dalam penelitian eksperimental komputasi, karena dilaksanakan melalui serangkaian eksperimen dengan menerapkan model *deep learning* pada lingkungan komputasi. Eksperimen dilakukan dengan mengonfigurasi dan melatih model YOLOv5 baseline serta model YOLOv5 yang menggunakan *backbone* EfficientNet-B4 untuk mendeteksi nilai mata uang, kemudian mengevaluasi kinerjanya berdasarkan hasil pengujian menggunakan dataset citra uang kertas.

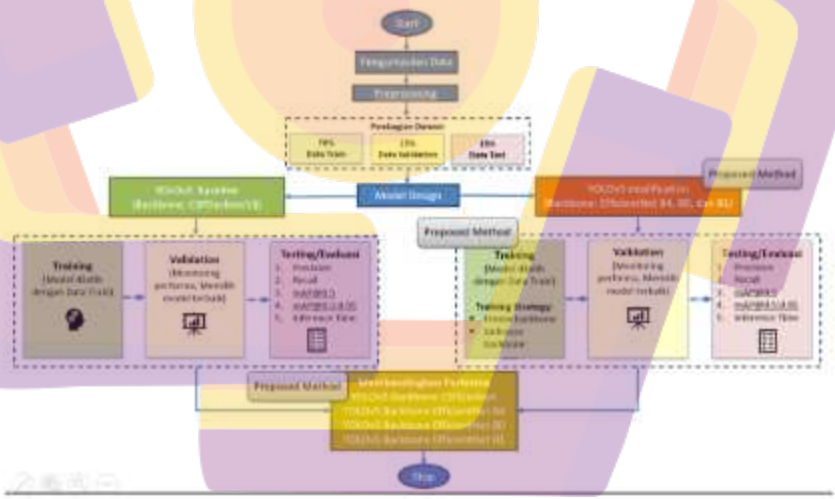
3.2 Metode Pengumpulan Data

Pengumpulan data menjelaskan tentang sumber data dan bagaimana proses pengumpulan data dilakukan. Sumber data pada penelitian ini berasal dari data sekunder, yaitu dataset publik yang telah tersedia melalui website https://roboflow.com.

3.3 Metode Analsis Data

Data yang telah dikumpulkan selanjutnya dianalisis untuk diolah menjadi informasi yang relevan dengan tujuan penelitian. Analisis data dilakukan dengan mengevaluasi kinerja sistem deteksi nilai mata uang yang dikembangkan, baik pada model YOLOv5 baseline maupun model YOLOv5 dengan *backbone* EfficientNet-B4. Proses analisis difokuskan pada perhitungan metrik evaluasi kinerja secara kuantitatif untuk mengukur tingkat akurasi deteksi dan ketelitian model dalam mendeteksi serta mengklasifikasikan nilai mata uang. Hasil analisis tersebut kemudian dibandingkan untuk mengetahui pengaruh penggunaan *backbone* EfficientNet-B4 terhadap kinerja sistem deteksi nilai mata uang.

3.4 Alur Penelitian



Gambar 3.1 Alur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan metode. Gambar 3.1 menunjukkan alur penelitian yang terdiri dari beberapa tahap antara lain:

a) Pengumpulan data

Dataset yang digunakan pada penelitian ini adalah citra uang kertas Rupiah yang masih berlaku dan beredar secara resmi di Indonesia saat ini, meliputi pecahan Rp1.000, Rp2.000, Rp5.000, Rp10.000, Rp20.000, Rp50.000, dan Rp100.000. Dataset pada penelitian ini terbagi menjadi 8 kelas yang terdiri dari 7 kelas nominal mata uang Rupiah serta 1 kelas tambahan berupa non-uang. Penambahan kelas non-uang bertujuan untuk membantu model membedakan objek uang dan objek selain uang sehingga kemampuan deteksi model menjadi lebih robust serta mengurangi potensi kesalahan deteksi pada kondisi nyata.

Dalam proses pengumpulan data, citra tidak hanya diambil pada kondisi ideal, tetapi juga mencakup berbagai kondisi yang umum ditemui dalam penggunaan sehari-hari. Variasi tersebut meliputi latar belakang yang kompleks, perbedaan intensitas pencahayaan (terang, redup, maupun *backlight*), variasi sudut dan orientasi pengambilan gambar, jarak pengambilan yang berbeda, serta kondisi fisik uang yang tidak ideal seperti terlipat, kusut, sebagian tertutup (*occlusion*), dan mengalami penurunan kualitas visual akibat penggunaan. Selain itu, dataset juga memuat kondisi di mana warna, tekstur, atau pola uang kertas memiliki kemiripan dengan latar belakang di sekitarnya, seperti dedaunan, bangunan, meja, maupun objek lainnya. Karakteristik ini meningkatkan tingkat kesulitan deteksi karena batas antara objek dan latar belakang menjadi kurang jelas. Beberapa citra juga diambil pada lingkungan luar ruangan dengan latar

belakang yang beragam sehingga model dilatih untuk mengenali uang kertas meskipun berada pada kondisi yang memiliki banyak gangguan visual (*background clutter*). Keberagaman kondisi tersebut bertujuan agar model dapat mempelajari karakteristik uang kertas secara lebih komprehensif sehingga memiliki kemampuan generalisasi yang lebih baik ketika diimplementasikan pada lingkungan nyata.

Gambar 3.2 Sampel Dataset Uang Kertas Rupiah

b) Pra-pemrosesan data (*preprocessing*)

Preprocessing atau pra-pemrosesan data bertujuan untuk mempersiapkan data agar lebih mudah digunakan dalam tahap pemodelan. Tahap preprocessing yang dilakukan pada penelitian ini yakni mengubah ukuran gambar (*resize*) untuk menstandarkan gambar dikarenakan teknik pengambilan gambar yang berbeda-beda. Seluruh gambar pada dataset disesuaikan ukurannya ke resolusi input yang digunakan oleh model YOLOv5 dan EfficientNet-B4 sebelum diproses

lebih lanjut, sehingga setiap gambar memiliki dimensi yang seragam dan sesuai dengan kebutuhan arsitektur jaringan. Tujuannya adalah memastikan bahwa input gambar sesuai dengan format yang diperlukan oleh YOLOv5 dan EfficientNet-B4. Kemudian tahap selanjutnya dilakukan Normalisasi dimana nilai piksel gambar (0-255) dinormalisasi ke skala 0-1. Hal ini dilakukan untuk membantu model memproses data lebih stabil dan efisien. Tahap berikutnya dilakukan Augmentasi data untuk meningkatkan variasi gambar sehingga model dapat belajar dari berbagai kondisi. Teknik augmentasi meliputi: Rotasi untuk memutar gambar, Flipping untuk membalik gambar secara horizontal atau vertical, dan Brightness/Contrast untuk mengubah tingkat kecerahan dan kontras gambar. Selanjutnya, dataset dibagi menjadi data latih, data validasi, dan data uji dengan proporsi masing-masing sebesar 70%, 15%, dan 15%. Pembagian dataset ini telah dilakukan sebelumnya melalui platform Roboflow, sehingga pada penelitian ini tidak melakukan proses pembagian dataset kembali untuk menjaga konsistensi dan validitas evaluasi model. Penelitian ini telah melakukan penambahan dataset secara keseluruhan yang mencakup data training, data validation, dan data testing dengan tetap mempertahankan skema pembagian 70:15:15 agar distribusi data tetap seimbang. Proporsi 15% pada data validasi telah cukup merepresentasikan distribusi data untuk memantau performa model selama training, melakukan *tuning* parameter, serta mendeteksi potensi *overfitting*. Penambahan jumlah data validasi secara berlebihan dapat mengurangi jumlah data latih yang digunakan untuk proses pembelajaran model, sehingga berpotensi menurunkan kemampuan model dalam

mempelajari karakteristik visual uang rupiah secara optimal. Selain itu, skema pembagian 70%:15%:15% ini merupakan konfigurasi yang umum digunakan pada penelitian *deep learning* dan *object detection* karena mampu memberikan evaluasi model yang stabil dan objektif. Rangkaian proses ini memastikan bahwa dataset siap digunakan untuk melatih model, sehingga model dapat mendeteksi uang kertas dengan lebih akurat di berbagai kondisi.

c) Pemodelan (*model design*)

Pemodelan (*model design*) pada penelitian ini menjelaskan proses pembangunan sistem deteksi nilai mata uang rupiah menggunakan arsitektur YOLOv5 dengan variasi *backbone*. Setelah tahap *preprocessing* data selesai, sistem deteksi dikembangkan dengan membangun dua model utama yaitu model YOLOv5 dengan *backbone* default CSPDarknet dan model YOLOv5 yang menggunakan arsitektur EfficientNet-B4 sebagai *backbone* pengganti. Penggunaan EfficientNet-B4 dimaksudkan untuk mengevaluasi pengaruh penggantian *backbone* terhadap kemampuan model dalam mengekstraksi fitur visual uang kertas.

d) Training

Tahap training merupakan proses utama dalam penelitian ini yang bertujuan untuk melatih model agar mampu mengenali dan mendeteksi nilai nominal uang kertas secara akurat. Training dilakukan menggunakan dua pendekatan model, yaitu YOLOv5 baseline dan YOLOv5 dengan *backbone* EfficientNet-B4, menggunakan konfigurasi dataset yang sama agar perbandingan performa bersifat adil.

Pada YOLOv5 baseline, seluruh komponen jaringan, termasuk *backbone, neck*, dan *head*, dilatih secara end-to-end sejak awal. Proses ini memungkinkan model mempelajari fitur visual dan pola objek secara langsung dari dataset uang kertas. Sedangkan pada model YOLOv5 dengan *backbone* EfficientNet-B4, training dilakukan menggunakan metode bertahap (*freeze* dan *unfreeze*). Konsep *freeze* dan *unfreeze* layer adalah bagian dari strategi *transfer learning* yaitu teknik di mana model yang sudah dilatih pada satu tugas digunakan kembali untuk tugas lain dan disesuaikan. Dengan melakukan *freezing* bagian tertentu dari pretrained model, proses pelatihan dapat menjadi lebih stabil serta berpotensi meningkatkan efisiensi komputasi. Strategi ini memungkinkan model mempertahankan fitur dasar yang telah dipelajari sebelumnya sehingga proses pelatihan menjadi lebih terarah dan dapat mempengaruhi performa serta efisiensi pelatihan model [14].

Pada tahap *freeze*, bobot EfficientNet-B4 dipertahankan tetap dan tidak diperbarui, sehingga proses pembelajaran difokuskan pada bagian *neck* dan *head* YOLOv5. Pendekatan ini bertujuan untuk menyesuaikan mekanisme deteksi terhadap fitur-fitur visual yang telah diekstraksi oleh *backbone* pretrained secara lebih stabil. Selanjutnya, pada tahap *unfreeze*, seluruh bobot jaringan, termasuk *backbone* EfficientNet-B4, dibuka kembali dan dilatih secara bersamaan. Tahap ini memungkinkan model melakukan penyesuaian lanjutan terhadap karakteristik spesifik dataset uang kertas rupiah, seperti variasi pencahayaan, kondisi fisik uang, dan latar belakang, dengan perubahan bobot yang lebih terkontrol. Dengan pendekatan training ini, model diharapkan

mampu mencapai keseimbangan antara stabilitas pembelajaran, akurasi deteksi, dan kemampuan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

e) Evaluasi

Evaluasi menjelaskan tentang evaluasi kinerja model yang digunakan untuk mengetahui seberapa baik model tersebut bekerja. Setelah pelatihan intensif menggunakan model YOLOv5 dengan CSPDarknet sebagai *backbone* default dan YOLOv5 dengan *backbone* EfficientNet-B4, kedua model tersebut dievaluasi menggunakan data uji guna menilai kinerjanya. Evaluasi ini bertujuan untuk melihat kemampuan model dalam mengenali dan mengklasifikasikan data yang belum pernah dipelajari, sehingga mencerminkan penerapannya pada kondisi nyata. Evaluasi performa kedua model ini dilakukan menggunakan metrik Precision, Recall, mAP@0.5, mAP@0.5:0.95, serta waktu inferensi untuk memberikan gambaran yang lebih menyeluruh mengenai kemampuan model dalam mendeteksi dan mengklasifikasikan objek. Hal ini memungkinkan penilaian kinerja model tidak hanya dari ketepatan deteksi, tetapi juga seberapa cepat model bekerja dalam mendeteksi objek. Selanjutnya perhitungan nilai *improvement* dilakukan untuk menganalisis pengaruh penggantian *backbone* EfficientNet-B4 terhadap kinerja YOLOv5. Melalui analisis ini, dapat diketahui apakah penggantian *backbone* EfficientNet-B4 memberikan peningkatan atau justru penurunan kinerja YOLOv5 dalam mendeteksi dan mengklasifikasikan objek.

BAB 4
HASIL PENELITIAN DAN PEMBAHASAN

4.1 Konfigurasi dan Skenario Pengujian

Pengembangan arsitektur model, pelatihan dan pengujian diimplementasikan dalam Google Colab Environment dengan menggunakan IPython Widgets digunakan sebagai modul Python utama dalam lingkungan pengembangan. Proses pembuatan dataset, pelabelan, serta pembagian data (*split balancing*) dilakukan menggunakan Roboflow, kemudian dataset diunduh dan disimpan pada Local SSD (*Solid State Drive*) sebagai media penyimpanan utama selama proses pelatihan dan evaluasi model. Penggunaan Local SSD dipilih karena memiliki kecepatan akses data yang lebih tinggi dan tidak bergantung pada koneksi jaringan, sehingga mampu mengurangi hambatan *input/output* yang sering terjadi saat menggunakan Google Drive. Dengan demikian, proses pelatihan menjadi lebih stabil, efisien, dan konsisten, terutama pada pelatihan model deteksi objek yang membutuhkan akses data secara intensif.

Dataset yang digunakan dalam penelitian ini terdiri dari 2.400 gambar uang kertas rupiah yang diambil dari berbagai sudut dan kondisi pencahayaan. Dataset ini telah dibagi menjadi menjadi tiga bagian yaitu data latih (*train*) sebanyak 1.680 gambar, data validasi (*validation*) sebanyak 360 gambar, dan data uji (*test*) sebanyak 360 gambar. Pembagian dataset ini bertujuan untuk memastikan proses pelatihan model berjalan optimal, sekaligus memungkinkan evaluasi performa model secara objektif menggunakan data yang tidak pernah dilihat sebelumnya.

Sebelum seluruh proses pelatihan dan pengujian dilakukan, terlebih dahulu dilakukan pengaturan terhadap konfigurasi parameter dan strategi pelatihan model untuk memastikan bahwa perbandingan kinerja antara YOLOv5 *baseline* dan YOLOv5 dengan *backbone* EfficientNet-B4 dilakukan secara objektif dan terkontrol. Rincian pengaturan parameter dan strategi pelatihan model yang digunakan dapat dilihat pada Tabel 4.1 berikut:

Tabel 4.1 Parameter dan Strategi Pelatihan

Model	Parameter	Value
YOLOv5 Baseline (CSPDarknet *Backbone*)	Image Size	640 × 640
	Batch Size	16
	Epoch	30
	Optimizer	SGD
	Learning Rate	0.01
	Device	GPU (CUDA)
YOLOv5 + EfficientNet-B4 (Stage A – Freeze *Backbone*)	*Backbone*	Freeze
	Image Size	640 × 640
	Batch Size	4
	Epoch	30
	Optimizer	AdamW
	Learning Rate	0.001
	Device	GPU (CUDA)
YOLOv5 + EfficientNet-B4 (Stage B – Unfreeze *Backbone*)	*Backbone*	Unfreeze
	Image Size	640 × 640
	Batch Size	4
	Epoch	30
	Optimizer	AdamW
	Learning Rate	0.0005
	Bobot Awal	Hasil Stage A
	Device	GPU (CUDA)

Pengujian dilakukan dalam tiga tahap, masing-masing dengan parameter atau kondisi tertentu:

1. Deteksi nilai mata uang dengan model YOLOv5 Default (CSPDarknet) sebagai baseline *backbone*.

2. Deteksi nilai mata uang dengan model YOLOv5 dengan arsitektur EfficientNet-B4 sebagai *backbone*.
3. Deteksi nilai mata uang dengan model YOLOv5 dengan arsitektur EfficientNet-B0 dan B1 sebagai *backbone*.

4.2 Hasil Training Model

Model YOLOv5 dengan *backbone* EfficientNet-B4 dilatih menggunakan pendekatan dua tahap (*two-stage training*), yaitu:

- Stage A: *Backbone* EfficientNet-B4 dibekukan (*freeze*) selama 30 epoch, menggunakan optimizer AdamW dan cosine learning rate scheduler. Tahap ini bertujuan untuk menyesuaikan head deteksi YOLO terhadap karakteristik fitur dari EfficientNet tanpa mengubah bobot *backbone* secara signifikan.
- Stage B: Seluruh layer dibuka (*unfreeze*) dan dilatih ulang selama 30 epoch menggunakan optimizer AdamW dengan cosine learning rate. Bobot awal pada tahap ini diambil dari hasil Stage A. Strategi ini bertujuan untuk menyempurnakan proses pembelajaran fitur secara menyeluruh.

Sedangkan Ukuran batch pada kedua tahap ditetapkan sebesar 4, menyesuaikan keterbatasan memori GPU akibat penggunaan *backbone* yang lebih kompleks. Adapun hasil training kedua model ditunjukkan Gambar 4.1 dan Gambar 4.2 sebagai berikut:

Gambar 4.1 Hasil Training YOLOv5 Baseline

Hasil training model YOLOv5 baseline menunjukkan bahwa model mampu mendeteksi sebagian besar kelas uang rupiah dengan performa yang cukup baik pada dataset yang terdiri dari 8 kelas. Dataset tersebut mencakup 7 kelas mata uang rupiah dan 1 kelas non-uang. Penambahan kelas non-uang menyebabkan tingkat kompleksitas deteksi meningkat karena model tidak hanya dituntut mengenali nominal uang rupiah, tetapi juga membedakan objek uang dan objek selain uang.

Berdasarkan hasil evaluasi, model memperoleh nilai precision sebesar 0.802, recall sebesar 0.668, mAP@0.5 sebesar 0.773, dan mAP@0.5:0.95 sebesar 0.561. Nilai precision sebesar 80,2% menunjukkan bahwa sebagian besar objek yang diprediksi oleh model merupakan objek yang benar, meskipun masih terdapat sejumlah kesalahan deteksi (*false positive*). Sementara itu, nilai recall sebesar 66,8% menunjukkan bahwa model telah mampu mendeteksi sebagian besar objek uang pada dataset, namun masih terdapat objek yang belum berhasil dikenali sehingga menghasilkan *false negative*. Nilai mAP@0.5 sebesar 77,3% menunjukkan bahwa model memiliki kemampuan deteksi yang cukup baik pada threshold IoU 0,5. Namun, nilai mAP@0.5:0.95 sebesar 56,1% yang lebih rendah menunjukkan bahwa ketepatan lokalisasi *bounding box* pada threshold IoU yang lebih ketat masih menjadi tantangan. Perbedaan yang cukup besar antara mAP@0.5 dan mAP@0.5:0.95 mengindikasikan bahwa meskipun objek umumnya berhasil

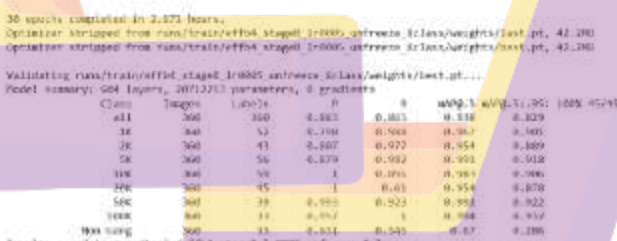
terdeteksi, posisi dan ukuran *bounding box* yang dihasilkan belum selalu presisi pada berbagai tingkat IoU. Kondisi tersebut sejalan dengan penelitian Diwan et al. yang menjelaskan bahwa *localization error* masih menjadi salah satu permasalahan utama pada algoritma YOLO, terutama ketika objek memiliki karakteristik visual yang mirip dan berada pada latar belakang yang kompleks [4]. Selain itu, Ramos dan Sappa juga menyatakan bahwa variasi skala objek, *occlusion*, perubahan posisi objek, serta *intra-class variation* masih menjadi tantangan utama dalam sistem *object detection* modern karena dapat memengaruhi kualitas ekstraksi fitur dan ketepatan lokalisasi objek. Pada penelitian ini, tantangan tersebut terlihat pada dataset yang mengandung variasi pencahayaan, latar belakang yang kompleks, kemiripan warna antara uang dan lingkungan sekitar, serta variasi orientasi dan kondisi fisik uang seperti terlipat atau sebagian tertutup, sehingga memengaruhi kemampuan model dalam menentukan lokasi objek secara presisi.

Pada masing-masing kelas uang rupiah, performa model menunjukkan hasil yang cukup baik meskipun terdapat variasi antar kelas. Kelas Rp100.000 memperoleh performa terbaik dengan nilai precision sebesar 0.863, recall sebesar 0.909, mAP@0.5 sebesar 0.955, dan mAP@0.5:0.95 sebesar 0.670. Hasil ini menunjukkan bahwa model mampu mengenali dan melokalisasi objek Rp100.000 dengan sangat baik. Kelas Rp5.000 juga menunjukkan performa yang tinggi dengan nilai mAP@0.5 sebesar 0.924 dan recall sebesar 0.875, yang mengindikasikan kemampuan deteksi yang baik pada berbagai kondisi citra. Sementara itu, kelas Rp10.000 dan Rp50.000 memperoleh nilai mAP@0.5 masing-masing sebesar 0.884 dan 0.867, yang menunjukkan kemampuan deteksi yang relatif stabil. Kelas

Rp20.000 memiliki nilai precision yang tinggi sebesar 0.879, namun recall yang lebih rendah yaitu 0.486, yang menunjukkan bahwa model cenderung menghasilkan prediksi yang benar ketika mendeteksi objek tersebut, tetapi masih gagal menemukan sebagian objek Rp20.000 yang terdapat pada dataset. Secara keseluruhan, sebagian besar kelas uang rupiah memperoleh nilai mAP@0.5 di atas 0,80, menunjukkan bahwa model cukup efektif dalam mempelajari karakteristik visual utama dari berbagai nominal uang rupiah.

Namun, performa yang berbeda terlihat pada kelas non-uang. Kelas ini memperoleh precision sebesar 1.0, recall sebesar 0.0, mAP@0.5 sebesar 0.266, dan mAP@0.5:0.95 sebesar 0.086. Kondisi tersebut menunjukkan bahwa model hampir tidak menghasilkan prediksi yang salah terhadap objek non-uang, tetapi pada saat yang sama gagal mendeteksi objek non-uang yang terdapat pada dataset pengujian. Nilai precision yang mencapai 1,000 bukan menunjukkan kemampuan deteksi yang sangat baik, melainkan mengindikasikan bahwa model sangat jarang atau bahkan tidak memberikan prediksi pada kelas non-uang sehingga tidak menghasilkan *false positive*. Sebaliknya, nilai recall sebesar 0,000 menunjukkan bahwa sebagian besar objek non-uang tidak berhasil terdeteksi oleh model. Hasil ini mengindikasikan bahwa YOLOv5 baseline masih mengalami kesulitan dalam mempelajari karakteristik kelas non-uang dibandingkan kelas uang rupiah. Kondisi tersebut dapat disebabkan oleh jumlah sampel non-uang yang relatif lebih sedikit dibandingkan kelas lainnya serta tingginya keragaman visual pada objek non-uang, sehingga model lebih fokus mempelajari karakteristik objek uang daripada membangun representasi fitur yang kuat untuk kelas non-uang.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa YOLOv5 baseline telah mampu mendeteksi sebagian besar nominal uang Rupiah dengan cukup baik, yang ditunjukkan oleh nilai mAP@0.5 sebesar 0.773 dan precision sebesar 0.802. Sebagian besar kelas uang juga memperoleh nilai mAP@0.5 di atas 0.80, menandakan bahwa model telah berhasil mempelajari karakteristik visual utama dari berbagai nominal uang Rupiah. Namun, performa yang rendah pada kelas non-uang menunjukkan bahwa keberadaan kelas tersebut meningkatkan kompleksitas proses deteksi karena model tidak hanya harus membedakan antar nominal uang, tetapi juga membedakan objek uang dan objek selain uang. Selain itu, selisih antara nilai mAP@0.5 (0.773) dan mAP@0.5:0.95 (0.561) menunjukkan bahwa ketepatan lokalisasi *bounding box* pada berbagai tingkat IoU masih dapat ditingkatkan. Hasil ini menjadi dasar yang penting untuk mengevaluasi pengaruh penggunaan backbone EfficientNet-B4 dalam meningkatkan kualitas ekstraksi fitur, ketepatan lokalisasi objek, serta kemampuan model dalam membedakan objek uang dan non-uang secara lebih akurat.



Gambar 4.2 Hasil Training YOLOv5 dengan *Backbone* EfficientNet-B4

Hasil training YOLOv5 dengan backbone EfficientNet-B4 menunjukkan peningkatan performa dibandingkan model YOLOv5 baseline. Berdasarkan hasil

evaluasi, model memperoleh nilai precision sebesar 0.883, recall sebesar 0.865, mAP@0.5 sebesar 0.938, dan mAP@0.5:0.95 sebesar 0.829. Nilai tersebut menunjukkan bahwa model mampu melakukan deteksi objek dengan tingkat ketepatan yang tinggi baik dalam proses klasifikasi maupun lokalisasi objek. Nilai mAP@0.5 sebesar 93.8% menunjukkan bahwa sebagian besar objek uang berhasil dideteksi dengan baik pada threshold IoU 0,5, sedangkan nilai mAP@0.5:0.95 sebesar 82.9% menunjukkan bahwa model juga mampu menghasilkan *bounding box* yang lebih presisi pada threshold IoU yang lebih ketat. Hasil ini mengindikasikan bahwa penggunaan EfficientNet-B4 sebagai backbone mampu meningkatkan kualitas ekstraksi fitur sehingga model lebih efektif dalam mengenali karakteristik visual uang kertas Rupiah pada berbagai kondisi pengambilan gambar.

Jika dibandingkan dengan YOLOv5 baseline, penggunaan backbone EfficientNet-B4 menghasilkan peningkatan performa yang signifikan pada seluruh metrik evaluasi. Nilai precision meningkat dari 0.802 menjadi 0.883, recall meningkat dari 0.668 menjadi 0.865, mAP@0.5 meningkat dari 0.773 menjadi 0.938, dan mAP@0.5:0.95 meningkat dari 0.561 menjadi 0.829. Peningkatan recall menunjukkan bahwa model mampu mendeteksi lebih banyak objek yang terdapat pada dataset sehingga jumlah objek yang tidak terdeteksi (*false negative*) menjadi lebih sedikit. Sementara itu, peningkatan nilai mAP@0.5 dan mAP@0.5:0.95 menunjukkan bahwa model tidak hanya lebih baik dalam mengenali objek, tetapi juga lebih akurat dalam menentukan lokasi objek melalui *bounding box* yang dihasilkan. Hasil ini mengindikasikan bahwa penggunaan EfficientNet-B4 sebagai backbone mampu menghasilkan representasi fitur yang lebih kaya sehingga

meningkatkan kemampuan model dalam mendeteksi berbagai nominal uang Rupiah pada kondisi visual yang beragam.

Peningkatan yang paling menonjol terlihat pada nilai mAP@0.5:0.95 yang meningkat dari 56,.1% pada YOLOv5 baseline menjadi 82.9% pada YOLOv5 dengan backbone EfficientNet-B4. Metrik mAP@0.5:0.95 merupakan indikator evaluasi yang lebih ketat karena mengukur performa deteksi pada berbagai threshold *Intersection over Union* (IoU) mulai dari 0.5 hingga 0.95. Oleh karena itu, metrik ini tidak hanya menilai keberhasilan model dalam mengenali objek, tetapi juga mengevaluasi ketepatan posisi dan ukuran *bounding box* yang dihasilkan. Peningkatan yang signifikan pada metrik ini menunjukkan bahwa backbone EfficientNet-B4 mampu menghasilkan representasi fitur yang lebih informatif sehingga proses lokalisasi objek menjadi lebih akurat dibandingkan backbone CSPDarknet pada YOLOv5 baseline. Dengan demikian, model tidak hanya lebih baik dalam mengenali nominal uang Rupiah, tetapi juga lebih presisi dalam menentukan lokasi objek pada citra meskipun terdapat variasi pencahayaan, orientasi, dan latar belakang yang kompleks.

Pada masing-masing kelas uang Rupiah, performa model menunjukkan hasil yang sangat baik dan relatif konsisten. Seluruh kelas uang memperoleh nilai mAP@0.5 di atas 0.95, dengan nilai tertinggi dicapai oleh kelas Rp100.000 sebesar 0.994 dan kelas Rp50.000 sebesar 0.992. Selain itu, sebagian besar kelas juga memperoleh nilai mAP@0.5:0.95 mendekati atau melebihi 0.90 serta nilai recall yang tinggi. Kelas Rp100.000 dan Rp10.000 bahkan mencapai nilai recall sebesar 1.00, menunjukkan bahwa seluruh objek pada kelas tersebut berhasil dideteksi oleh

model. Hasil ini menunjukkan bahwa EfficientNet-B4 mampu menghasilkan representasi fitur yang lebih kaya dan diskriminatif sehingga model dapat membedakan karakteristik visual antar nominal uang Rupiah dengan lebih baik. Kemampuan tersebut tetap terjaga meskipun dataset mengandung variasi pencahayaan, orientasi pengambilan gambar, latar belakang yang kompleks, serta kemiripan warna dan pola visual antar nominal uang.

Performa kelas non-uang juga mengalami peningkatan dibandingkan YOLOv5 baseline. Setelah menggunakan backbone EfficientNet-B4, kelas non-uang memperoleh precision sebesar 0.631, recall sebesar 0.545, mAP@0.5 sebesar 0.670, dan mAP@0.5:0.95 sebesar 0.286. Meskipun nilai precision lebih rendah dibandingkan model baseline, peningkatan recall dari 0.0 menjadi 0.545 menunjukkan bahwa model telah mampu mendeteksi sebagian besar objek non-uang yang sebelumnya gagal dikenali. Selain itu, nilai mAP@0.5 meningkat dari 0.266 menjadi 0.670, sedangkan mAP@0.5:0.95 meningkat dari 0.086 menjadi 0.286, yang menunjukkan adanya peningkatan kemampuan model dalam mengenali dan melokalisasi objek non-uang. Hasil ini mengindikasikan bahwa penggunaan EfficientNet-B4 mampu menghasilkan fitur yang lebih representatif sehingga model lebih efektif dalam membedakan objek uang dan non-uang, meskipun jumlah data non-uang relatif sedikit dan memiliki variasi visual yang lebih beragam dibandingkan kelas uang Rupiah.

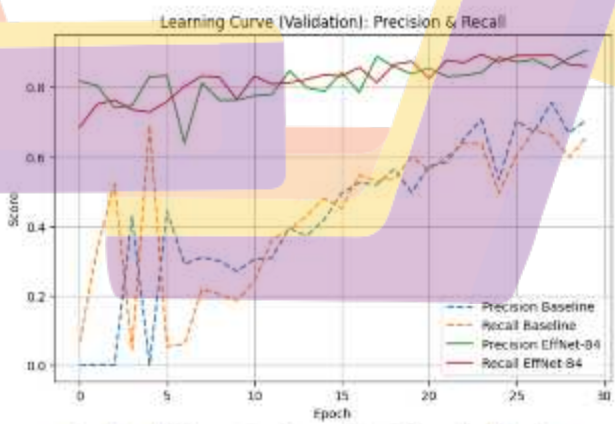
Secara keseluruhan, hasil evaluasi menunjukkan bahwa penggantian backbone CSPDarknet dengan EfficientNet-B4 memberikan pengaruh yang signifikan terhadap peningkatan performa YOLOv5. Peningkatan terlihat pada

seluruh metrik evaluasi, yaitu precision, recall, mAP@0.5, dan mAP@0.5:0.95. Hasil tersebut menunjukkan bahwa EfficientNet-B4 mampu menghasilkan representasi fitur yang lebih kaya dan diskriminatif sehingga model lebih efektif dalam mengenali karakteristik visual berbagai nominal uang Rupiah serta membedakannya dari objek non-uang. Selain meningkatkan kemampuan klasifikasi, penggunaan EfficientNet-B4 juga memberikan peningkatan pada ketepatan lokalisasi objek yang ditunjukkan oleh kenaikan nilai mAP@0.5:0.95 dari 56,1% menjadi 82,9%. Hasil ini memperkuat bahwa backbone memiliki peran yang sangat penting dalam menentukan kualitas deteksi objek karena berfungsi sebagai ekstraktor fitur utama pada jaringan. Temuan ini sejalan dengan konsep *compound scaling* yang diperkenalkan pada EfficientNet, yaitu peningkatan kedalaman (*depth*), lebar (*width*), dan resolusi (*resolution*) jaringan secara seimbang sehingga kemampuan ekstraksi fitur menjadi lebih optimal dibandingkan backbone CSPDarknet pada model baseline.

Dalam konteks deteksi nilai mata uang Rupiah, kemampuan ekstraksi fitur yang baik sangat penting karena setiap pecahan uang memiliki karakteristik visual yang berbeda, seperti warna dominan, gambar pahlawan nasional, pola ornamen, angka nominal, serta elemen grafis lainnya. Selain itu, beberapa pecahan uang memiliki kemiripan warna, tekstur, maupun pola visual yang dapat meningkatkan risiko kesalahan klasifikasi. Tantangan deteksi juga menjadi lebih kompleks karena dataset yang digunakan mengandung variasi pencahayaan, latar belakang yang beragam, perubahan orientasi objek, kondisi uang yang terlipat atau sebagian tertutup, serta objek non-uang yang memiliki karakteristik visual yang berbeda-

beda. Dengan kemampuan ekstraksi fitur yang lebih kuat, EfficientNet-B4 mampu menangkap informasi visual yang lebih detail dan representatif sehingga model dapat membedakan antar nominal uang maupun antara objek uang dan non-uang dengan lebih baik. Kemampuan tersebut berkontribusi terhadap peningkatan performa deteksi yang ditunjukkan oleh kenaikan nilai precision, recall, mAP@0.5, dan mAP@0.5:0.95 dibandingkan model YOLOv5 baseline.

Analisis dinamika proses pelatihan model tidak hanya dilakukan pada hasil evaluasi akhir, tetapi juga pada perkembangan metrik kinerja selama pelatihan. *Learning curve* ditunjukkan pada Gambar 4.3, digunakan untuk menggambarkan perubahan nilai precision dan recall terhadap jumlah epoch sehingga memungkinkan pengamatan terhadap pola pembelajaran, stabilitas prediksi, dan proses konvergensi model. Pada penelitian ini, *learning curve* dianalisis untuk membandingkan perilaku pembelajaran antara model YOLOv5 baseline dan model YOLOv5 dengan *backbone* EfficientNet-B4 sebagai dasar dalam menjelaskan perbedaan karakteristik kinerja kedua model.



Gambar 4.3 Learning Kurva Precission dan Recall

Grafik *Learning Curve (Validation): Precision & Recall* menunjukkan perkembangan performa model selama proses training pada data validasi, khususnya dalam aspek precision dan recall antara YOLOv5 baseline dan YOLOv5 dengan backbone EfficientNet-B4. Grafik ini penting untuk melihat stabilitas proses pembelajaran model, kemampuan generalisasi, serta efektivitas backbone dalam meningkatkan kualitas deteksi objek.

Pada grafik terlihat bahwa kurva EfficientNet-B4 memiliki pola yang jauh lebih stabil dibandingkan YOLOv5 baseline. Sejak epoch awal, nilai precision EfficientNet-B4 telah berada pada tingkat yang relatif tinggi, yaitu berkisar antara 0.75 hingga 0.90, dan cenderung stabil hingga akhir proses training. Sebaliknya, precision pada YOLOv5 baseline menunjukkan fluktuasi yang cukup besar, terutama pada epoch-epoch awal, sebelum akhirnya meningkat secara bertahap. Stabilitas kurva precision pada backbone EfficientNet-B4 menunjukkan bahwa model mampu menghasilkan prediksi yang lebih konsisten dengan tingkat *false positive* yang relatif rendah. Kondisi ini mengindikasikan bahwa proses ekstraksi fitur yang dilakukan oleh backbone EfficientNet-B4 mampu menghasilkan representasi fitur yang lebih baik sehingga model dapat mempelajari karakteristik objek uang Rupiah dengan lebih cepat dan efektif sejak tahap awal pelatihan.

Sementara itu, recall EfficientNet-B4 juga menunjukkan pola yang relatif stabil dengan kecenderungan meningkat selama proses training. Pada epoch awal, nilai recall berada pada kisaran 0.70, kemudian meningkat secara bertahap hingga mencapai sekitar 0.86–0.90 pada akhir training. Kondisi ini menunjukkan bahwa model semakin mampu mendeteksi lebih banyak objek yang terdapat pada dataset

seiring bertambahnya jumlah epoch. Dalam konteks sistem deteksi nilai mata uang, peningkatan recall menunjukkan bahwa model semakin sedikit melewatkan objek uang Rupiah yang terdapat pada citra sehingga kemampuan deteksi nominal uang menjadi lebih baik. Hal ini penting karena pada sistem deteksi nilai mata uang, objek yang gagal terdeteksi (*false negative*) dapat menyebabkan nominal uang tidak terbaca oleh sistem. Kurva recall yang relatif stabil juga menunjukkan bahwa model mampu mempelajari karakteristik visual masing-masing pecahan uang Rupiah secara konsisten meskipun terdapat variasi posisi, pencahayaan, sudut pengambilan gambar, kondisi uang yang tidak ideal, serta latar belakang yang beragam. Selain itu, stabilitas kurva tersebut mengindikasikan bahwa proses training berlangsung secara lebih konvergen dan terkontrol sehingga model memiliki kemampuan generalisasi yang lebih baik terhadap data yang tidak pernah dilihat sebelumnya.

Berbeda dengan EfficientNet-B4, YOLOv5 baseline menunjukkan pola pembelajaran yang lebih fluktuatif selama proses training. Nilai precision mengalami perubahan yang cukup besar pada epoch-epoch awal, bahkan pada beberapa epoch berada pada nilai yang sangat rendah sebelum meningkat secara bertahap seiring bertambahnya jumlah epoch. Kondisi ini menunjukkan bahwa model baseline membutuhkan waktu yang lebih lama untuk mempelajari karakteristik visual objek secara optimal. Fluktuasi yang terjadi pada tahap awal pelatihan juga mengindikasikan bahwa proses pembelajaran model belum stabil sehingga masih menghasilkan kesalahan prediksi (*false positive*) yang relatif tinggi. Meskipun performa precision terus mengalami peningkatan pada epoch-epoch berikutnya, nilai yang dicapai hingga akhir training masih berada di bawah model

dengan backbone EfficientNet-B4. Hasil ini menunjukkan bahwa kemampuan ekstraksi fitur pada backbone CSPDarknet belum seefektif EfficientNet-B4 dalam menghasilkan representasi fitur yang konsisten untuk mendukung proses deteksi objek.

Recall YOLOv5 baseline juga menunjukkan fluktuasi yang cukup besar selama proses training, terutama pada epoch-epoch awal. Nilai recall sempat mengalami peningkatan dan penurunan yang tajam sebelum akhirnya menunjukkan tren peningkatan yang lebih stabil pada epoch berikutnya. Pada akhir training, nilai recall mencapai sekitar 0,65–0,68, yang menunjukkan bahwa model telah mampu mendeteksi sebagian besar objek pada dataset, meskipun masih terdapat sejumlah objek yang belum berhasil dikenali. Pola ini mengindikasikan bahwa kemampuan model baseline dalam mendeteksi seluruh objek pada dataset belum sepenuhnya stabil selama proses pembelajaran. Fluktuasi tersebut dapat disebabkan oleh keterbatasan backbone CSPDarknet dalam menghasilkan representasi fitur yang konsisten pada dataset yang memiliki variasi pencahayaan, orientasi, latar belakang yang kompleks, serta kondisi objek yang beragam. Dibandingkan dengan EfficientNet-B4 yang mampu mempertahankan nilai recall tinggi secara konsisten, YOLOv5 baseline memerlukan lebih banyak epoch untuk mencapai performa yang stabil sehingga menunjukkan proses konvergensi yang lebih lambat.

Jika dibandingkan secara keseluruhan, backbone EfficientNet-B4 menunjukkan proses konvergensi yang lebih cepat dan stabil dibandingkan YOLOv5 baseline. Hal ini terlihat dari nilai precision EfficientNet-B4 yang relatif tinggi dan stabil sejak awal training, nilai recall yang meningkat secara konsisten,

serta tingkat fluktuasi yang lebih kecil dibandingkan model baseline. Kondisi tersebut menunjukkan bahwa backbone EfficientNet-B4 mampu menghasilkan representasi fitur yang lebih baik sehingga model lebih mudah membedakan objek uang dan non-uang pada berbagai kondisi visual.

Selain itu, stabilitas kurva menunjukkan bahwa strategi *transfer learning* dengan pendekatan *freeze–unfreeze* membantu proses pembelajaran berlangsung secara lebih terkontrol. Pada tahap *freeze*, bobot backbone hasil *pretraining* dipertahankan sehingga model dapat memanfaatkan fitur-fitur umum yang telah dipelajari sebelumnya, seperti tepi objek, tekstur, bentuk, pola visual, dan detail ornamen uang kertas. Kondisi ini membuat proses pembelajaran awal menjadi lebih stabil serta membantu model mengenali karakteristik visual dasar dari masing-masing nominal uang Rupiah. Selanjutnya, pada tahap *unfreeze*, seluruh lapisan backbone dilatih kembali sehingga model dapat menyesuaikan fitur yang telah dipelajari dengan karakteristik spesifik dataset penelitian. Proses ini memungkinkan model mempelajari perbedaan warna, pola, angka nominal, tekstur, serta ciri visual lainnya yang membedakan setiap pecahan uang Rupiah. Hasil evaluasi menunjukkan bahwa pendekatan tersebut berkontribusi terhadap peningkatan kemampuan deteksi dan lokalisasi objek, yang ditunjukkan oleh peningkatan nilai recall dari 0.668 menjadi 0.865, mAP@0.5 dari 0.773 menjadi 0.938, serta mAP@0.5:0.95 dari 0.561 menjadi 0.829 dibandingkan YOLOv5 baseline. Dengan demikian, kombinasi backbone EfficientNet-B4 dan strategi *freeze–unfreeze* tidak hanya meningkatkan stabilitas proses training, tetapi juga menghasilkan kemampuan deteksi dan lokalisasi objek yang lebih akurat pada

berbagai kondisi citra uang Rupiah, termasuk variasi pencahayaan, orientasi objek, latar belakang yang kompleks, dan kondisi uang yang tidak ideal.

Peningkatan kemampuan deteksi ini berkaitan erat dengan proses ekstraksi fitur pada jaringan yang digunakan dalam model deteksi objek. Pada arsitektur YOLO, backbone berfungsi sebagai komponen utama yang mengekstraksi fitur visual dari citra masukan sebelum fitur tersebut diteruskan ke bagian *neck* dan *head* untuk proses deteksi dan klasifikasi objek. Mahasin dan Dewi dalam penelitiannya menjelaskan bahwa pemilihan backbone pada arsitektur YOLO sangat mempengaruhi kualitas fitur yang dihasilkan sehingga dapat berdampak langsung pada performa deteksi objek [12]. Hasil penelitian ini mendukung pernyataan tersebut, di mana penggantian backbone CSPDarknet pada YOLOv5 dengan EfficientNet-B4 menghasilkan peningkatan pada seluruh metrik evaluasi, yaitu precision, recall, mAP@0.5, dan mAP@0.5:0.95. Peningkatan tersebut menunjukkan bahwa EfficientNet-B4 mampu menghasilkan representasi fitur yang lebih informatif dan diskriminatif sehingga model lebih efektif dalam mengenali karakteristik visual berbagai nominal uang Rupiah serta menentukan lokasi objek secara lebih akurat.

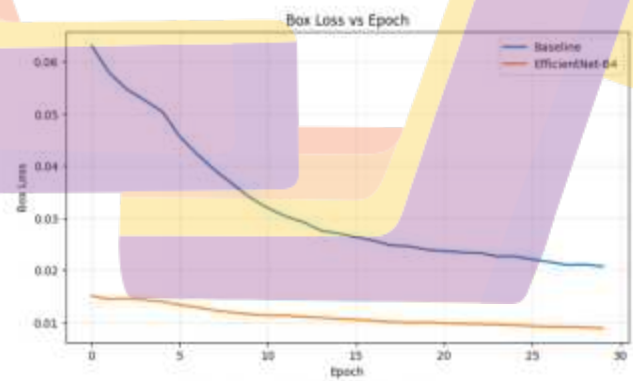
Dalam penelitian ini, penggunaan EfficientNet-B4 sebagai backbone memberikan kemampuan ekstraksi fitur yang lebih baik dibandingkan backbone CSPDarknet pada YOLOv5 baseline. Tan dan Le menjelaskan bahwa EfficientNet menggunakan pendekatan *compound scaling*, yaitu peningkatan kedalaman (*depth*), lebar (*width*), dan resolusi (*resolution*) jaringan secara seimbang sehingga mampu menghasilkan representasi fitur yang lebih kaya dan efisien [19]. Dengan

representasi fitur yang lebih baik, model dapat mengenali pola visual objek secara lebih akurat pada berbagai kondisi citra. Temuan tersebut sejalan dengan hasil penelitian ini yang menunjukkan peningkatan pada seluruh metrik evaluasi setelah backbone diganti menggunakan EfficientNet-B4. Peningkatan nilai precision, recall, mAP@0.5, dan mAP@0.5:0.95 menunjukkan bahwa model mampu mengenali karakteristik visual uang Rupiah dengan lebih baik sekaligus menghasilkan lokalisasi objek yang lebih akurat. Kemampuan ini menjadi penting karena dataset yang digunakan mengandung variasi pencahayaan, orientasi pengambilan gambar, latar belakang yang kompleks, kondisi uang yang tidak ideal, serta objek non-uang yang memiliki karakteristik visual beragam. Temuan ini juga didukung oleh penelitian Li et al. yang menunjukkan bahwa integrasi EfficientNet dengan model YOLOv5 dapat meningkatkan kemampuan model dalam mengenali objek karena *feature extractor* yang lebih kuat mampu menangkap detail visual objek secara lebih efektif [15]. Pada penelitian ini, kemampuan tersebut tercermin dari peningkatan performa deteksi pada hampir seluruh kelas uang Rupiah serta meningkatnya kemampuan model dalam membedakan objek uang dan non-uang dibandingkan YOLOv5 baseline.

Secara keseluruhan, pola *learning curve* ini menunjukkan bahwa *backbone* EfficientNet-B4 membantu model YOLOv5 mencapai stabilitas pembelajaran lebih cepat dan mempertahankan kinerja deteksi yang lebih konsisten sepanjang proses pelatihan. Hal ini terlihat dari kurva yang cenderung stabil setelah beberapa epoch awal, yang menandakan bahwa model mampu menyesuaikan parameter jaringan secara lebih efektif selama proses pelatihan. Stabilitas tersebut sejalan dengan

penerapan strategi *freeze* dan *unfreeze* dalam pendekatan *transfer learning*. Dobrzycki et al menjelaskan bahwa pembekuan lapisan tertentu pada model *pretrained* dapat meningkatkan kestabilan pelatihan dan mempercepat proses konvergensi dengan mengurangi fluktuasi gradien pada tahap awal pembelajaran. Selain itu, pendekatan ini juga dilaporkan mampu menurunkan penggunaan memori GPU hingga 28% dibandingkan metode *full fine-tuning*, sehingga proses pelatihan menjadi lebih efisien tanpa mengorbankan performa model [14].

Dalam melengkapi analisis hasil pelatihan, dilakukan pengamatan terhadap perilaku kurva loss selama proses training seperti ditunjukkan oleh kurva box loss, objectness loss, dan classification loss disajikan untuk menggambarkan dinamika pembelajaran model dalam mengoptimalkan lokalisasi objek, pendeteksian keberadaan objek, dan klasifikasi kelas seiring bertambahnya epoch. Analisis kurva loss ini memberikan konteks tambahan dalam menafsirkan kinerja model, khususnya terkait kestabilan proses pelatihan dan kecenderungan konvergensi yang dicapai oleh masing-masing arsitektur *backbone*.



Gambar 4.4 Kurva Box Loss

Kurva *Box Loss* menggambarkan perkembangan kemampuan model dalam melokalisasi *bounding box*, baik dari sisi posisi maupun ukuran kotak deteksi, selama proses pelatihan. Nilai *box loss* yang semakin kecil menunjukkan bahwa selisih antara *bounding box* prediksi dan *ground truth* semakin kecil, sehingga posisi dan ukuran kotak deteksi yang dihasilkan model semakin akurat dan mendekati objek yang sebenarnya di dalam citra. Artinya Semakin kecil nilai *box loss*, maka semakin baik kemampuan model dalam melakukan lokalisasi objek secara presisi.

Berdasarkan grafik, kedua model sama-sama menunjukkan tren penurunan *box loss* seiring bertambahnya epoch. Hal ini menunjukkan bahwa proses pembelajaran berjalan dengan baik karena model semakin mampu menyesuaikan prediksi *bounding box* dengan posisi objek sebenarnya pada dataset. Namun, penurunan *box loss* pada EfficientNet-B4 terlihat jauh lebih stabil dan konsisten dibandingkan YOLOv5 baseline.

Pada YOLOv5 baseline, nilai *box loss* pada awal training berada di kisaran 0,063 kemudian menurun secara bertahap hingga mencapai sekitar 0,021 pada akhir training. Penurunan yang cukup signifikan pada epoch-epoch awal menunjukkan bahwa model mulai mempelajari posisi dan ukuran objek uang Rupiah dengan semakin baik seiring bertambahnya proses pelatihan. Namun, nilai *box loss* YOLOv5 baseline tetap lebih tinggi dibandingkan model dengan backbone EfficientNet-B4 sepanjang proses training. Kondisi tersebut menunjukkan bahwa backbone CSPDarknet masih memiliki keterbatasan dalam menghasilkan representasi fitur yang cukup detail untuk menentukan lokasi objek secara presisi.

Keterbatasan ini menjadi lebih terlihat pada dataset yang memiliki variasi pencahayaan, orientasi pengambilan gambar, latar belakang yang kompleks, kondisi uang yang tidak ideal, serta kemiripan warna dan pola visual antar nominal uang Rupiah. Akibatnya, proses penyesuaian posisi *bounding box* memerlukan waktu lebih lama dan menghasilkan tingkat kesalahan lokalisasi yang lebih tinggi dibandingkan EfficientNet-B4.

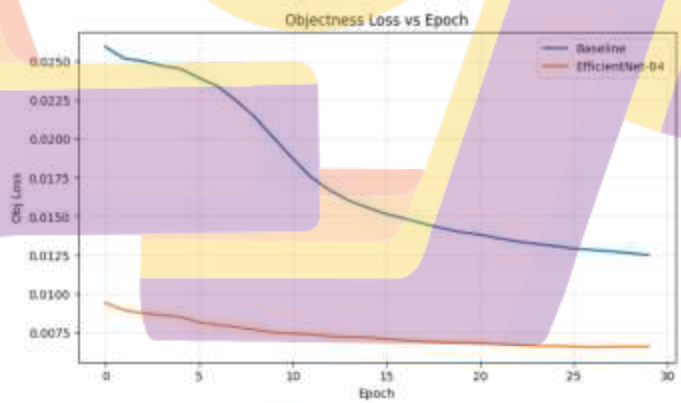
Meskipun nilai *box loss* terus menurun hingga akhir training, kurva YOLOv5 baseline masih berada jauh di atas kurva EfficientNet-B4. Hasil ini sejalan dengan nilai mAP@0.5:0.95 yang diperoleh pada tahap training, di mana model YOLOv5 baseline hanya mencapai 56.1%, sedangkan model dengan backbone EfficientNet-B4 mencapai 82.9%. Dengan demikian, nilai *box loss* yang lebih tinggi pada YOLOv5 baseline menunjukkan bahwa kemampuan lokalisasi objek masih lebih rendah dibandingkan model yang menggunakan EfficientNet-B4 sebagai backbone.

Sementara itu, EfficientNet-B4 menunjukkan nilai *box loss* yang jauh lebih rendah sejak awal training, yaitu sekitar 0.015 dan terus menurun secara konsisten hingga mencapai sekitar 0,009 pada akhir epoch. Nilai *box loss* yang rendah menunjukkan bahwa model mampu menghasilkan prediksi *bounding box* yang lebih dekat dengan posisi objek sebenarnya. Stabilitas kurva juga menunjukkan bahwa proses pembelajaran berlangsung lebih konvergen dan terkontrol tanpa fluktuasi yang signifikan.

Dalam konteks deteksi nilai mata uang rupiah, hasil ini menunjukkan bahwa EfficientNet-B4 memiliki kemampuan ekstraksi fitur yang lebih baik sehingga

model lebih mudah mengenali detail visual uang seperti angka nominal, pola ornamen, warna, dan tekstur pada berbagai kondisi citra. Kemampuan tersebut membantu model menentukan lokasi objek uang secara lebih akurat sehingga kualitas lokalisasi *bounding box* meningkat. Hal ini juga sejalan dengan peningkatan nilai mAP@0.5 dan terutama mAP@0.5:0.95 pada EfficientNet-B4 dibandingkan YOLOv5 baseline, karena metrik tersebut sangat dipengaruhi oleh ketepatan posisi *bounding box*.

Selain itu, rendahnya *box loss* pada EfficientNet-B4 juga menunjukkan bahwa strategi *transfer learning* dan *freeze–unfreeze* membantu backbone mempertahankan fitur umum hasil *pretraining* sehingga proses pembelajaran lokalisasi objek menjadi lebih stabil sejak awal training. Dengan demikian, penggunaan backbone EfficientNet-B4 tidak hanya meningkatkan kemampuan klasifikasi nominal uang, tetapi juga meningkatkan kualitas lokalisasi objek secara keseluruhan pada sistem deteksi nilai mata uang rupiah.



Gambar 4.5 Kurva Objectness Loss

Kurva *Objectness Loss* (obj loss) merepresentasikan kemampuan model dalam memprediksi keberadaan objek pada setiap kandidat *bounding box*, yaitu membedakan antara area yang benar-benar mengandung objek dan area latar belakang. Nilai *objectness loss* yang semakin kecil menunjukkan bahwa model semakin mampu mengenali apakah suatu lokasi pada citra mengandung objek atau hanya merupakan latar belakang, sehingga kesalahan deteksi objek palsu (*false positive*) dapat dikurangi. Semakin kecil nilai *objectness loss*, maka semakin baik kemampuan model dalam membedakan area yang benar-benar berisi objek dengan latar belakang (*background*).

Berdasarkan grafik, kedua model menunjukkan tren penurunan *objectness loss* seiring bertambahnya epoch. Hal ini menunjukkan bahwa model semakin mampu mengenali keberadaan objek pada citra selama proses training berlangsung. Namun, EfficientNet-B4 menunjukkan nilai *objectness loss* yang jauh lebih rendah dan lebih stabil dibandingkan YOLOv5 baseline.

Pada YOLOv5 baseline, nilai *objectness loss* pada awal training berada di sekitar 0.026 kemudian menurun secara bertahap hingga mencapai sekitar 0.013 pada akhir training. Penurunan ini menunjukkan bahwa model semakin mampu membedakan area yang mengandung objek uang dan area latar belakang (*background*) selama proses pembelajaran berlangsung. Namun, laju penurunan kurva terlihat lebih lambat dan nilai *objectness loss* tetap lebih tinggi dibandingkan EfficientNet-B4 sepanjang proses training. Kondisi tersebut menunjukkan bahwa backbone CSPDarknet masih memiliki keterbatasan dalam menghasilkan representasi fitur yang cukup kuat untuk membedakan objek uang dan non-objek

secara optimal, terutama pada dataset yang memiliki variasi pencahayaan, orientasi pengambilan gambar, latar belakang yang kompleks, serta kemiripan warna dan pola visual antar objek. Akibatnya, model berpotensi mengalami kesalahan dalam mengidentifikasi keberadaan objek, baik berupa *false positive* maupun *false negative*. Kondisi ini juga tercermin pada hasil evaluasi, di mana YOLOv5 baseline memperoleh nilai recall sebesar 0.668 yang lebih rendah dibandingkan model dengan backbone EfficientNet-B4.

Dalam sistem deteksi nilai mata uang Rupiah, kemampuan mengenali keberadaan objek secara akurat sangat penting karena kesalahan dalam tahap identifikasi objek dapat memengaruhi proses klasifikasi nominal dan lokalisasi *bounding box*. Temuan ini sejalan dengan penelitian Xu et al. yang menjelaskan bahwa YOLOv5 merupakan algoritma *object detection* yang memiliki performa baik pada berbagai aplikasi, namun masih menghadapi tantangan dalam mendeteksi objek pada kondisi visual yang kompleks, termasuk latar belakang yang beragam dan variasi karakteristik objek, yang dapat memengaruhi kualitas deteksi dan ketepatan lokalisasi objek.

Sementara itu, model YOLOv5 dengan backbone EfficientNet-B4 menunjukkan nilai *objectness loss* yang jauh lebih rendah sejak awal training, yaitu sekitar 0,009 dan terus menurun secara konsisten hingga mencapai sekitar 0,0065 pada akhir training. Nilai *objectness loss* yang rendah menunjukkan bahwa model lebih cepat mempelajari karakteristik area yang mengandung objek uang Rupiah dan mampu membedakannya dari area latar belakang dengan lebih baik. Kondisi ini mengindikasikan bahwa model memiliki kemampuan yang lebih baik dalam

menentukan keberadaan objek pada citra sehingga proses deteksi menjadi lebih efektif.

Selain itu, kurva *objectness loss* EfficientNet-B4 menunjukkan pola yang relatif halus dan stabil tanpa fluktuasi yang berarti selama proses training. Stabilitas tersebut menunjukkan bahwa proses pembelajaran berlangsung secara lebih konvergen dan terkontrol, sehingga model mampu menyesuaikan parameter jaringan secara efektif pada setiap epoch. Kemampuan ini mengindikasikan bahwa fitur yang diekstraksi oleh EfficientNet-B4 lebih representatif dalam membedakan objek uang, objek non-uang, dan latar belakang.

Hasil tersebut sejalan dengan peningkatan performa selama training yang ditunjukkan oleh nilai recall sebesar 0.865 dan mAP@0.5 sebesar 0.938, yang lebih tinggi dibandingkan YOLOv5 baseline. Dengan demikian, rendahnya nilai *objectness loss* pada EfficientNet-B4 menunjukkan bahwa model memiliki kemampuan yang lebih baik dalam mengenali keberadaan objek uang Rupiah pada berbagai kondisi citra, termasuk variasi pencahayaan, orientasi pengambilan gambar, latar belakang yang kompleks, serta keberadaan objek non-uang.

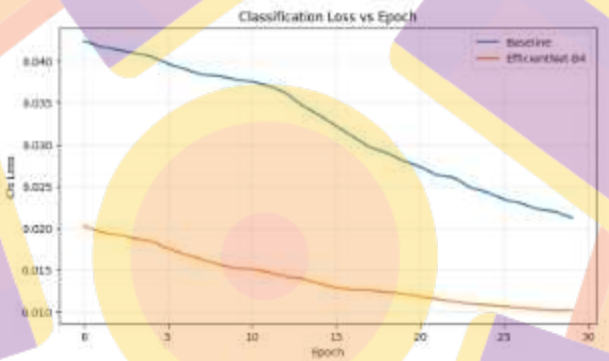
Dalam konteks deteksi nilai mata uang rupiah, hasil ini menunjukkan bahwa EfficientNet-B4 mampu menghasilkan fitur yang lebih representatif untuk membedakan objek uang dan non-uang pada berbagai kondisi citra. Hal ini penting karena uang rupiah sering memiliki warna dan pola yang mirip dengan latar belakang tertentu atau objek lain pada citra. Kemampuan backbone dalam mengenali area objek secara lebih akurat membantu model mengurangi kesalahan prediksi pada area *background* sehingga *false positive* menjadi lebih rendah.

Selain itu, rendahnya *objectness loss* pada EfficientNet-B4 juga berkaitan dengan peningkatan nilai precision dan recall yang diperoleh model. Ketika model mampu mengenali area objek secara lebih baik, maka probabilitas objek uang terdeteksi dengan benar menjadi lebih tinggi dan kesalahan deteksi objek palsu dapat dikurangi. Hal ini sejalan dengan hasil evaluasi yang menunjukkan bahwa EfficientNet-B4 menghasilkan performa mAP@0.5 dan mAP@0.5:0.95 yang lebih tinggi dibandingkan YOLOv5 baseline.

Stabilitas kurva pada EfficientNet-B4 juga menunjukkan bahwa strategi *transfer learning* dan *freeze–unfreeze* membantu model mempertahankan fitur umum hasil *pretraining* sehingga proses identifikasi keberadaan objek berlangsung lebih stabil sejak awal training. Dengan demikian, penggunaan backbone EfficientNet-B4 tidak hanya meningkatkan kemampuan lokalisasi objek, tetapi juga meningkatkan kemampuan model dalam mengenali keberadaan objek uang rupiah secara lebih akurat pada berbagai kondisi visual.

Dalam konteks penelitian ini, kemampuan membedakan objek dan latar belakang sangat penting karena objek yang dideteksi adalah uang kertas rupiah yang sering muncul pada berbagai kondisi latar belakang, seperti meja, tangan, atau permukaan lainnya. Jika model tidak mampu membedakan objek dengan latar belakang secara baik, maka sistem dapat menghasilkan deteksi yang tidak akurat, misalnya menganggap pola latar sebagai bagian dari uang kertas atau sebaliknya tidak mengenali keberadaan uang dalam citra. Kemampuan model dalam membedakan objek dan latar belakang sangat dipengaruhi oleh kualitas *feature extraction* pada jaringan. Tan dan Le menjelaskan bahwa EfficientNet

meningkatkan kualitas representasi fitur melalui pendekatan *compound scaling*, yaitu peningkatan kedalaman jaringan, lebar jaringan, dan resolusi input secara seimbang. Representasi fitur yang lebih kaya memungkinkan model menangkap pola visual objek secara lebih detail dan jelas, sehingga membantu model dalam membedakan area yang mengandung objek dengan area latar belakang. Dengan demikian, proses prediksi keberadaan objek pada setiap kandidat *bounding box* dapat dilakukan dengan lebih akurat.



Gambar 4.6 Kurva *Classification Loss*

Kurva *Classification Loss* (cls loss) menunjukkan perkembangan kemampuan model dalam membedakan dan mengklasifikasikan kelas objek yang terdapat pada citra. Dalam penelitian ini, kelas objek yang dimaksud adalah nominal uang kertas rupiah. Nilai *classification loss* yang semakin kecil menunjukkan bahwa model semakin mampu mempelajari perbedaan karakteristik visual antar kelas sehingga kesalahan dalam memprediksi kelas objek dapat diminimalkan. Artinya semakin kecil nilai *classification loss*, maka semakin baik kemampuan model dalam membedakan masing-masing kelas objek pada dataset.

Berdasarkan grafik, kedua model menunjukkan tren penurunan *classification loss* seiring bertambahnya epoch. Hal ini menunjukkan bahwa proses pembelajaran klasifikasi berjalan dengan baik karena model semakin mampu mengenali perbedaan antar kelas uang rupiah dan kelas non-uang. Namun, EfficientNet-B4 menunjukkan nilai *classification loss* yang jauh lebih rendah dan lebih stabil dibandingkan YOLOv5 baseline.

Pada YOLOv5 baseline, nilai *classification loss* pada awal training berada di sekitar 0.042 kemudian menurun secara bertahap hingga mencapai sekitar 0.021 pada akhir training. Penurunan ini menunjukkan bahwa model semakin mampu mempelajari karakteristik visual masing-masing kelas objek selama proses pelatihan berlangsung. Dengan bertambahnya jumlah epoch, model menjadi lebih baik dalam membedakan berbagai nominal uang Rupiah serta objek non-uang yang terdapat pada dataset.

Meskipun demikian, nilai *classification loss* YOLOv5 baseline tetap relatif lebih tinggi dibandingkan model yang menggunakan backbone EfficientNet-B4 sepanjang proses training. Kondisi ini mengindikasikan bahwa backbone CSPDarknet belum mampu menghasilkan representasi fitur yang seoptimal EfficientNet-B4 dalam membedakan karakteristik visual antar kelas. Tantangan tersebut menjadi lebih besar karena beberapa nominal uang Rupiah memiliki kemiripan warna, pola ornamen, tekstur, maupun elemen visual lainnya yang dapat meningkatkan risiko kesalahan klasifikasi.

Nilai *classification loss* yang lebih tinggi menunjukkan bahwa proses pemisahan antar kelas masih belum sepenuhnya optimal. Akibatnya, model

berpotensi menghasilkan kesalahan klasifikasi pada beberapa objek yang memiliki karakteristik visual yang mirip. Temuan ini sejalan dengan performa selama training yang menunjukkan bahwa performa deteksi YOLOv5 baseline masih berada di bawah model dengan backbone EfficientNet-B4, baik dari sisi precision, recall, maupun nilai mAP. Dengan demikian, nilai *classification loss* yang lebih tinggi pada YOLOv5 baseline mengindikasikan bahwa kemampuan model dalam membedakan antar kelas objek masih lebih rendah dibandingkan model yang menggunakan EfficientNet-B4 sebagai backbone.

Sementara itu, EfficientNet-B4 menunjukkan nilai *classification loss* yang jauh lebih rendah sejak awal training, yaitu sekitar 0.020 dan terus menurun secara stabil hingga mendekati 0.010 pada akhir epoch. Nilai *classification loss* yang lebih rendah menunjukkan bahwa model mampu mempelajari karakteristik visual masing-masing kelas secara lebih efektif sehingga proses klasifikasi antar nominal uang Rupiah dan kelas non-uang menjadi lebih akurat.

Selain itu, kurva *classification loss* EfficientNet-B4 menunjukkan pola penurunan yang relatif halus dan stabil tanpa fluktuasi yang berarti selama proses training. Kondisi ini mengindikasikan bahwa proses pembelajaran berlangsung secara lebih konvergen dan terkontrol, sehingga model mampu membangun representasi fitur yang lebih diskriminatif untuk membedakan karakteristik visual setiap kelas. Kemampuan tersebut sangat penting pada dataset yang digunakan karena beberapa nominal uang memiliki kemiripan warna, pola ornamen, dan tekstur, sementara kelas non-uang memiliki variasi visual yang sangat beragam.

Dalam konteks deteksi nilai mata uang Rupiah, hasil ini menunjukkan bahwa EfficientNet-B4 memiliki kemampuan ekstraksi fitur yang lebih baik dalam mengenali karakteristik visual setiap pecahan uang, seperti angka nominal, gambar pahlawan, warna dominan, pola ornamen, tekstur, serta elemen grafis khas lainnya. Kemampuan tersebut memungkinkan model membangun representasi fitur yang lebih diskriminatif sehingga proses klasifikasi antar kelas dapat dilakukan dengan lebih akurat. Selain itu, backbone EfficientNet-B4 juga menunjukkan kemampuan yang lebih baik dalam membedakan objek uang dan kelas non-uang yang memiliki karakteristik visual lebih beragam. Hal ini menjadi penting karena dataset yang digunakan tidak hanya terdiri atas berbagai nominal uang Rupiah, tetapi juga mencakup objek non-uang serta kondisi citra yang bervariasi. Dengan representasi fitur yang lebih kaya, model mampu membedakan antar nominal uang maupun membedakan uang dan non-uang secara lebih efektif meskipun terdapat kemiripan warna, pola visual, dan tekstur antar objek.

Kemampuan tersebut turut didukung oleh nilai *classification loss* yang lebih rendah dan lebih stabil dibandingkan YOLOv5 baseline sepanjang proses training. Hasil ini menunjukkan bahwa EfficientNet-B4 mampu mempelajari karakteristik visual setiap kelas secara lebih optimal sehingga menghasilkan peningkatan performa klasifikasi yang tercermin pada kenaikan nilai precision, recall, dan mAP. Dengan demikian, penggunaan EfficientNet-B4 memberikan kontribusi yang signifikan terhadap kemampuan model dalam mengenali dan mengklasifikasikan nilai mata uang Rupiah pada berbagai kondisi citra, termasuk variasi pencahayaan,

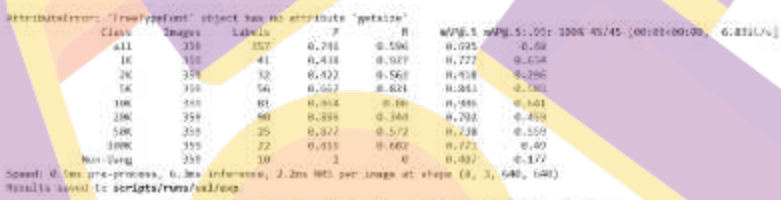
orientasi objek, sudut pengambilan gambar, latar belakang yang kompleks, serta kondisi uang yang tidak ideal.

Rendahnya *classification loss* pada EfficientNet-B4 juga berkaitan dengan peningkatan nilai precision dan recall yang diperoleh model. Ketika model mampu mengklasifikasikan objek dengan lebih baik, maka kesalahan prediksi kelas dapat dikurangi sehingga jumlah *false positive* dan *false negative* menjadi lebih rendah. Hal ini sejalan dengan hasil performa model selama training yang menunjukkan peningkatan nilai precision dari 0.802 menjadi 0.883, recall dari 0.668 menjadi 0.865, mAP@0.5 dari 0.773 menjadi 0.938 serta mAP@0.5:0.95 dari 0.561 menjadi 0.829 dibandingkan YOLOv5 baseline. Hasil ini menunjukkan bahwa EfficientNet-B4 mampu menghasilkan fitur yang lebih representatif sehingga model dapat membedakan antar kelas objek dengan lebih baik dan menghasilkan prediksi yang lebih akurat.

Kemampuan model dalam membedakan kelas objek ini sangat dipengaruhi oleh kualitas *feature extraction* yang dihasilkan oleh backbone. Dalam penelitian ini, pola kurva *classification loss* menunjukkan bahwa penggunaan backbone EfficientNet-B4 memberikan kontribusi positif terhadap kemampuan model dalam mengklasifikasikan nominal uang rupiah. Representasi fitur yang lebih diskriminatif memungkinkan model menangkap perbedaan karakteristik visual antar pecahan uang dan non uang dengan lebih jelas, sehingga kesalahan klasifikasi dapat dikurangi. Dengan demikian, sistem deteksi nilai mata uang berbasis *deep learning* yang diusulkan menjadi lebih andal dalam mengenali berbagai nominal uang rupiah secara akurat.

4.3 Hasil Evaluasi Model

Hasil training digunakan untuk menentukan apakah model telah mempelajari pola data secara memadai sebelum dilakukan evaluasi lebih lanjut pada data uji. Evaluasi ini dilakukan guna menilai kinerja masing-masing model dalam mendeteksi dan mengklasifikasikan nilai nominal uang kertas pada data uji yang belum pernah digunakan selama proses pelatihan. Hasil evaluasi kedua model ditunjukkan sebagai berikut:



Gambar 4.7 Hasil Evaluasi YOLOv5 Baseline

Hasil evaluasi YOLOv5 baseline menunjukkan bahwa model mampu melakukan deteksi objek uang Rupiah dengan performa yang cukup baik pada dataset 8 kelas yang terdiri dari 7 kelas mata uang Rupiah dan 1 kelas non-uang. Berdasarkan hasil evaluasi, model memperoleh nilai precision sebesar 0.746, recall sebesar 0.596, mAP@0.5 sebesar 0.695, dan mAP@0.5:0.95 sebesar 0.480.

Nilai precision sebesar 0.746 menunjukkan bahwa sebagian besar objek yang diprediksi model merupakan prediksi yang benar, meskipun masih terdapat sejumlah *false positive*. Hal ini menunjukkan bahwa YOLOv5 baseline telah mampu membedakan objek uang dan latar belakang pada sebagian besar kondisi citra, namun masih terdapat beberapa kesalahan deteksi. Sementara itu, nilai recall sebesar 0.596 menunjukkan bahwa model mampu mendeteksi sebagian objek uang pada dataset, tetapi masih terdapat cukup banyak objek yang gagal terdeteksi (*false*

negative). Dalam konteks deteksi nilai mata uang Rupiah, kondisi ini menunjukkan bahwa masih terdapat beberapa nominal uang yang belum berhasil dikenali model pada kondisi tertentu, seperti variasi pencahayaan, posisi objek yang miring, kondisi uang yang tidak ideal, atau latar belakang yang kompleks.

Nilai mAP@0.5 sebesar 69.5% menunjukkan bahwa model memiliki kemampuan deteksi objek yang cukup baik pada threshold IoU 0.5. Artinya, sebagian besar *bounding box* yang dihasilkan model telah mampu mengidentifikasi objek uang pada citra dengan tingkat kesesuaian yang memadai. Namun, ketika evaluasi dilakukan menggunakan threshold yang lebih ketat melalui mAP@0.5:0.95, nilainya menurun menjadi 48.0%. Penurunan ini menunjukkan bahwa kualitas lokalisasi *bounding box* pada YOLOv5 baseline pada dataset dengan kondisi tidak ideal, masih belum optimal pada berbagai tingkat IoU yang lebih tinggi. Dengan kata lain, model telah mampu mengenali keberadaan uang Rupiah, tetapi ketepatan posisi dan ukuran *bounding box* masih perlu ditingkatkan pada beberapa kondisi citra. Hal ini sejalan dengan penelitian Diwan et al. yang menjelaskan bahwa *localization error* masih menjadi salah satu tantangan utama pada algoritma YOLO, terutama ketika objek memiliki karakteristik visual yang mirip dan kondisi latar belakang yang kompleks. Selain itu, Ramos dan Sappa juga menyatakan bahwa variasi skala objek, *occlusion*, perubahan posisi objek, serta *intra-class variation* masih menjadi tantangan pada sistem *object detection* modern karena dapat memengaruhi kualitas ekstraksi fitur dan ketepatan lokalisasi objek.

Pada masing-masing kelas uang Rupiah, performa model menunjukkan variasi yang cukup besar. Kelas Rp10.000 memperoleh hasil terbaik dengan

precision sebesar 0.864, recall sebesar 0.860, mAP@0.5 sebesar 0.906, dan mAP@0.5:0.95 sebesar 0.641. Hasil ini menunjukkan bahwa model mampu mengenali sekaligus melokalisasi objek Rp10.000 dengan cukup baik dibandingkan kelas lainnya. Kelas Rp5.000 juga menunjukkan performa yang baik dengan nilai mAP@0.5 sebesar 0.841, sedangkan kelas Rp1.000 memperoleh nilai mAP@0.5 sebesar 0.777 dengan recall yang tinggi sebesar 0.927. Sementara itu, kelas Rp20.000 memiliki precision yang tinggi sebesar 0.886, namun recall yang relatif rendah yaitu 0,344, yang menunjukkan bahwa masih terdapat banyak objek Rp20.000 yang belum berhasil terdeteksi. Secara keseluruhan, sebagian besar kelas uang Rupiah memperoleh nilai mAP@0.5 di atas 0.70, yang menunjukkan bahwa YOLOv5 baseline cukup mampu mempelajari karakteristik visual utama dari berbagai nominal uang Rupiah.

Namun, performa yang berbeda terlihat pada kelas non-uang. Kelas ini memperoleh precision sebesar 1.0, recall sebesar 0.0, mAP@0.5 sebesar 0.407, dan mAP@0.5:0.95 sebesar 0.177. Kondisi ini menunjukkan bahwa model hampir tidak melakukan prediksi salah terhadap objek non-uang, tetapi pada saat yang sama gagal mendeteksi objek non-uang yang terdapat pada dataset pengujian. Nilai precision yang tinggi terjadi karena model sangat jarang atau bahkan tidak memberikan prediksi pada kelas non-uang sehingga tidak menghasilkan *false positive*. Salah satu penyebab utama kondisi tersebut adalah jumlah data non-uang yang relatif sedikit dibandingkan kelas lainnya sehingga model lebih dominan mempelajari karakteristik visual uang Rupiah. Selain itu, objek non-uang memiliki variasi bentuk, warna, tekstur, dan pola visual yang jauh lebih beragam

dibandingkan uang Rupiah sehingga proses ekstraksi fitur menjadi lebih sulit dan kemampuan deteksi kelas non-uang menjadi lebih rendah dibandingkan kelas lainnya.

Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa YOLOv5 baseline telah mampu mendeteksi nominal uang rupiah dengan cukup baik, namun masih memiliki keterbatasan pada kualitas lokalisasi *bounding box* dan kemampuan membedakan objek non-uang. Hasil ini menjadi dasar penting untuk membandingkan pengaruh penggunaan backbone EfficientNet-B4 terhadap peningkatan kualitas ekstraksi fitur, stabilitas deteksi, serta ketepatan lokalisasi objek pada sistem deteksi nilai mata uang rupiah.

Selain metrik akurasi, hasil evaluasi juga menunjukkan kinerja model dari sisi kecepatan pemrosesan. Waktu *inference* tercatat sekitar 6,3 ms per gambar, yang menunjukkan bahwa model mampu melakukan proses deteksi objek dengan cepat. Hasil ini menunjukkan bahwa YOLOv5 baseline memiliki potensi untuk diterapkan pada sistem deteksi nilai mata uang secara *real-time*, karena proses pengenalan nominal uang dan penentuan posisi objek dapat dilakukan dalam waktu yang relatif singkat dengan tetap mempertahankan performa deteksi yang cukup baik. Hal ini sejalan dengan penelitian Redmon et al yang menjelaskan bahwa YOLO melakukan deteksi objek secara langsung dalam satu jaringan dengan memprediksi lokasi *bounding box* dan kelas objek secara bersamaan, sehingga menghasilkan sistem deteksi yang cepat dan efisien [2].

Proses deteksi objek juga melibatkan tahap *Non-Maximum Suppression* (NMS) dengan waktu sekitar 2,2 ms per gambar. Nilai NMS ini menunjukkan

bahwa YOLOv5 baseline mampu melakukan proses penyaringan dan pemilihan *bounding box* objek secara cukup cepat untuk mengurangi prediksi yang tumpang tindih (*overlapping bounding box*). Hal ini penting dalam sistem deteksi nilai mata uang karena membantu model menentukan lokasi objek uang yang paling tepat sehingga hasil deteksi menjadi lebih akurat dan tidak menghasilkan banyak prediksi ganda pada objek yang sama.

Untuk mengevaluasi pengaruh penggantian *backbone* terhadap performa deteksi objek, maka model YOLOv5 dengan backbone EfficientNet-B4 kemudian diuji menggunakan dataset yang sama digunakan model YOLOv5 baseline. Hasil evaluasi model tersebut disajikan pada Gambar 4.8 di bawah ini.

Gambar 4.8 Hasil Evaluasi YOLOv5 dengan *Backbone* EfficientNet-B4

Hasil evaluasi YOLOv5 dengan backbone EfficientNet-B4 menunjukkan peningkatan performa yang cukup signifikan dibandingkan YOLOv5 baseline. Berdasarkan hasil evaluasi, model memperoleh nilai precision sebesar 0.774, recall sebesar 0.847, mAP@0.5 sebesar 0.901, dan mAP@0.5:0.95 sebesar 0.746.

Nilai recall yang mencapai 0.847 menunjukkan bahwa model mampu mendeteksi sebagian besar objek uang yang terdapat pada dataset sehingga jumlah objek yang gagal terdeteksi (*false negative*) menjadi lebih sedikit dibandingkan YOLOv5 baseline yang hanya memperoleh recall sebesar 0.596. Dalam konteks

deteksi nilai mata uang rupiah, peningkatan recall ini menunjukkan bahwa model semakin mampu mengenali berbagai nominal uang pada berbagai kondisi citra seperti perubahan pencahayaan, rotasi, sudut pengambilan gambar, maupun latar belakang yang kompleks.

Selain itu, nilai mAP@0.5 sebesar 90.1% menunjukkan bahwa model memiliki kemampuan deteksi objek yang sangat baik pada threshold IoU 0.5. Sementara itu, nilai mAP@0.5:0.95 sebesar 74.6% menunjukkan bahwa kualitas lokalisasi *bounding box* pada threshold yang lebih ketat juga mengalami peningkatan yang signifikan dibandingkan YOLOv5 baseline. Peningkatan ini menunjukkan bahwa backbone EfficientNet-B4 mampu menghasilkan representasi fitur yang lebih kaya dan detail sehingga model dapat menentukan posisi objek uang secara lebih presisi.

Pada masing-masing kelas uang Rupiah, performa model menunjukkan hasil yang baik dan relatif stabil. Kelas Rp50.000 memperoleh performa terbaik dengan precision sebesar 0.958, recall sebesar 0.911, mAP@0.5 sebesar 0.986, dan mAP@0.5:0.95 sebesar 0.897. Kelas Rp10.000 juga menunjukkan performa yang sangat baik dengan mAP@0.5 sebesar 0.968 dan mAP@0.5:0.95 sebesar 0.882. Selain itu, kelas Rp5.000 memperoleh mAP@0.5 sebesar 0.949, sedangkan kelas Rp100.000 mencapai mAP@0.5 sebesar 0.928. Hasil ini menunjukkan bahwa EfficientNet-B4 mampu mempelajari karakteristik visual utama pada masing-masing nominal uang Rupiah secara lebih efektif sehingga menghasilkan kemampuan deteksi dan klasifikasi yang lebih baik dibandingkan model baseline.

Pada kelas non-uang, model memperoleh precision sebesar 0.65, recall sebesar 0.9, mAP@0.5 sebesar 0.932, dan mAP@0.5:0.95 sebesar 0.304. Hasil ini menunjukkan bahwa EfficientNet-B4 memiliki kemampuan yang lebih baik dalam membedakan objek uang dan non-uang dibandingkan YOLOv5 baseline. Peningkatan recall pada kelas non-uang dari 0.0 pada model baseline menjadi 0.9 menunjukkan bahwa model mampu mengenali sebagian besar objek non-uang yang terdapat pada dataset pengujian. Selain itu, peningkatan nilai mAP@0.5 dari 0.407 menjadi 0.932 menunjukkan adanya peningkatan yang signifikan dalam kemampuan deteksi kelas non-uang. Hasil ini mengindikasikan bahwa EfficientNet-B4 mampu menghasilkan representasi fitur yang lebih baik sehingga model dapat membedakan objek uang dan non-uang secara lebih efektif meskipun jumlah data non-uang relatif sedikit dan memiliki variasi visual yang lebih kompleks.

Selain metrik akurasi, kinerja model juga dapat dilihat dari waktu pemrosesan. Hasil evaluasi menunjukkan bahwa waktu inference sekitar 40.8 ms per gambar, sedangkan proses Non-Maximum Suppression (NMS) memerlukan waktu sekitar 1.8 ms per gambar. Hal ini menunjukkan bahwa model mampu melakukan proses deteksi objek dengan efisien, meskipun menggunakan *backbone* yang lebih kompleks dibandingkan model baseline. Proses NMS yang relatif cepat menunjukkan bahwa penyaringan *bounding box* yang saling tumpang tindih dapat dilakukan secara efektif tanpa memberikan beban komputasi yang signifikan terhadap keseluruhan proses deteksi.

Hasil penelitian ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa penggunaan backbone yang lebih kuat dapat meningkatkan performa model deteksi objek. Mahasin dan Dewi menunjukkan bahwa pemilihan backbone pada arsitektur YOLO dapat mempengaruhi kualitas fitur yang dihasilkan oleh jaringan sehingga berdampak langsung pada akurasi deteksi objek [12]. Selain itu, Nugroho dan Baihaqi juga menegaskan bahwa penggantian backbone pada YOLOv5 dapat meningkatkan kemampuan model dalam mengekstraksi fitur visual sehingga membantu model membedakan kelas objek dengan lebih baik [3].

Pada Tabel 4.2 menunjukkan hasil perbandingan kinerja antara model YOLOv5 baseline dan YOLOv5 dengan *backbone* EfficientNet-B4 pada data uji. Dengan menggunakan rumus *percentage improvement* (*% Improvement*), penelitian ini bertujuan untuk mengetahui besarnya perubahan kinerja model setelah dilakukan penggantian *backbone*, baik berupa peningkatan maupun penurunan performa, pada metrik evaluasi Precision, Recall, mAP@0.5, dan mAP@0.5:0.95 sebagai berikut:

Tabel. 4.2 Model Evaluation Report

Metrik	YOLOv5 baseline	YOLOv5 + EfficientNet-B4	% Improvement
Precision	74.6	77.4	3.75
Recall	59.6	84.7	42.11
mAP@0.5	69.5	90.1	29.6
mAP@0.5:0.95	48.0	74.6	55.41
Inference Time	6.3 ms	40.8 ms	547

Tabel 4.2 di atas menunjukkan bahwa penggunaan backbone EfficientNet-B4 berhasil meningkatkan kemampuan deteksi dan kualitas lokalisasi objek dibandingkan YOLOv5 baseline. Model menjadi lebih mampu mendeteksi objek

uang rupiah pada berbagai kondisi citra khususnya kondisi citra yang tidak ideal dan menghasilkan *bounding box* yang lebih presisi, terutama pada evaluasi dengan *threshold* IoU yang lebih ketat. Hasil ini menunjukkan bahwa EfficientNet-B4 memiliki kemampuan ekstraksi fitur yang lebih baik dalam mengenali detail visual uang rupiah dan membedakan objek secara lebih akurat. Meskipun terjadi sedikit penurunan precision, peningkatan recall dan mAP menunjukkan bahwa performa deteksi model secara keseluruhan menjadi lebih baik. Dari sisi komputasi, peningkatan performa tersebut diikuti dengan waktu *inference* yang lebih tinggi karena kompleksitas backbone yang lebih besar.

Temuan ini sejalan dengan penelitian Li et al, dalam penelitian inspeksi pipa bawah laut juga menunjukkan bahwa penggunaan EfficientNet sebagai *backbone* pada YOLOv5 menghasilkan peningkatan akurasi dan stabilitas deteksi dibandingkan YOLOv5 baseline [13]. Dengan demikian, hasil penelitian ini memperkuat temuan sebelumnya bahwa *backbone* dengan kapasitas representasi yang lebih tinggi mampu meningkatkan performa deteksi, khususnya dalam tugas yang menuntut ketelitian lokalisasi objek yang tinggi.

Selain itu peningkatan waktu eksekusi dari 6.3 ms pada YOLOv5 baseline menjadi 40.8 ms pada YOLOv5 dengan backbone EfficientNet-B4 terjadi karena meningkatnya kompleksitas komputasi model. Berdasarkan penelitian *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks* [19], EfficientNet-B4 memiliki sekitar 19 juta parameter dan membutuhkan 4,2 miliar FLOPs (*Floating Point Operations*) untuk memproses satu citra. Jumlah parameter dan FLOPs yang lebih besar menyebabkan proses ekstraksi fitur memerlukan lebih

banyak operasi konvolusi, penggunaan memori yang lebih tinggi, serta waktu komputasi yang lebih lama dibandingkan backbone CSPDarknet pada YOLOv5 baseline. Kompleksitas tersebut berasal dari penerapan *compound scaling* yang meningkatkan kedalaman (*depth*), lebar (*width*), dan resolusi (*resolution*) jaringan secara bersamaan sehingga representasi fitur yang dihasilkan menjadi lebih kaya dan detail. Dampaknya terlihat pada hasil evaluasi Tabel 4.2, dimana penggunaan EfficientNet-B4 mampu meningkatkan Recall dari 59.6% menjadi 84.7%, mAP@0.5 dari 69.5% menjadi 90.1%, dan mAP@0.5:0.95 dari 48.0% menjadi 74.6%. Peningkatan mAP@0.5:0.95 sebesar 55.41% menunjukkan bahwa model mampu menghasilkan lokalisasi objek dan *bounding box* yang lebih presisi pada berbagai *threshold* IoU. Namun, peningkatan kualitas deteksi tersebut harus dibayar dengan bertambahnya waktu inferensi sekitar 547%, yang menunjukkan adanya *trade-off* antara akurasi dan kecepatan inferensi. Dengan kata lain, EfficientNet-B4 membutuhkan komputasi yang lebih besar sehingga waktu deteksi menjadi lebih lama, tetapi mampu menghasilkan kualitas ekstraksi fitur dan ketepatan lokalisasi objek yang lebih baik dibandingkan YOLOv5 baseline.

Dari sisi proses pelatihan, kinerja model juga dipengaruhi oleh keterbatasan jumlah data pelatihan dan durasi training. EfficientNet-B4 merupakan *backbone* dengan tingkat kompleksitas yang tinggi dan jumlah parameter yang besar, sehingga memerlukan dataset yang cukup banyak serta jumlah epoch pelatihan yang memadai agar pembelajaran fitur dapat mencapai kondisi konvergensi yang optimal. Dalam penelitian ini, jumlah epoch pelatihan dibatasi hingga 30 epoch sebagai konsekuensi dari keterbatasan sumber daya GPU. Pembatasan ini

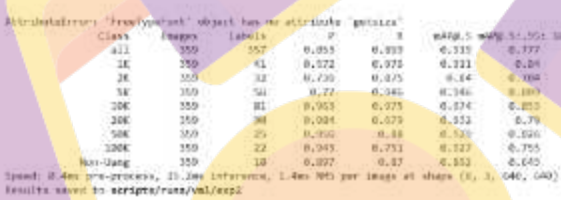
berpotensi menyebabkan proses optimasi belum sepenuhnya mencapai performa maksimal, sehingga masih terdapat peluang peningkatan akurasi deteksi apabila pelatihan dilakukan dengan durasi yang lebih panjang atau dukungan komputasi yang lebih besar.

Penelitian ini bertujuan untuk mengevaluasi secara eksperimental penggunaan arsitektur EfficientNet-B4 sebagai pengganti *backbone* default YOLOv5 (CSPDarknet) dalam sistem deteksi nilai mata uang rupiah. Evaluasi dilakukan untuk meninjau pengaruh penggantian *backbone* terhadap kinerja deteksi objek, yang meliputi kemampuan deteksi, ketelitian lokalisasi *bounding box*, serta efisiensi komputasi model.

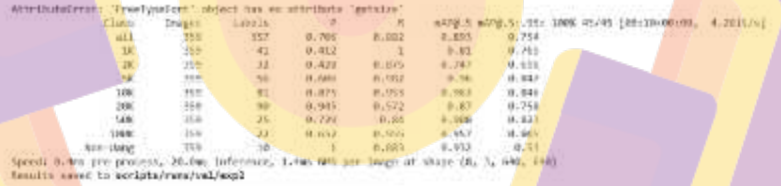
Berdasarkan hasil evaluasi penggunaan *backbone* EfficientNet-B4, penelitian ini selanjutnya tidak hanya berfokus pada *backbone* dengan kompleksitas tinggi, tetapi juga melibatkan EfficientNet-B0 dan EfficientNet-B1 sebagai pembanding dengan kompleksitas yang lebih ringan. Kedua *backbone* tersebut memiliki jumlah parameter yang lebih sedikit dan kebutuhan komputasi yang lebih rendah, sehingga memungkinkan proses pelatihan yang lebih efisien pada kondisi keterbatasan data dan sumber daya komputasi.

Pendekatan ini bertujuan untuk mengidentifikasi konfigurasi *backbone* yang paling optimal dalam mencapai keseimbangan antara akurasi deteksi dan efisiensi komputasi, khususnya pada sistem deteksi nilai mata uang rupiah. Dengan membandingkan *backbone* dengan tingkat kompleksitas yang berbeda, penelitian ini dapat menganalisis sejauh mana peningkatan kapasitas representasi fitur berkontribusi terhadap peningkatan kinerja deteksi, serta bagaimana dampaknya

terhadap waktu inferensi dan kebutuhan sumber daya GPU. Selain itu, pendekatan ini juga memungkinkan evaluasi apakah *backbone* dengan kompleksitas yang lebih ringan, seperti EfficientNet-B0 dan EfficientNet-B1, mampu memberikan performa yang kompetitif dengan beban komputasi yang lebih rendah, sehingga lebih sesuai untuk diimplementasikan pada lingkungan komputasi dengan keterbatasan sumber daya. Berikut adalah hasil evaluasi YOLOv5 dengan *Backbone* EfficientNet-B0 dan EfficientNet-B1.



Gambar 4.9 Hasil Evaluasi YOLOv5 dengan *Backbone* EfficientNet-B0



Gambar 4.10 Hasil Evaluasi YOLOv5 dengan *Backbone* EfficientNet-B1

Guna memperoleh gambaran yang lebih komprehensif mengenai pengaruh tingkat kompleksitas *backbone* terhadap kinerja YOLOv5, dilakukan perbandingan antara beberapa varian EfficientNet dengan kapasitas yang berbeda, yaitu EfficientNet-B0, EfficientNet-B1, dan EfficientNet-B4. Ketiga *backbone* tersebut dipilih untuk merepresentasikan *backbone* ringan, menengah, dan kompleks, sehingga memungkinkan analisis yang lebih menyeluruh terhadap hubungan antara kapasitas representasi fitur, kinerja deteksi, dan efisiensi komputasi. Hasil evaluasi

masing-masing konfigurasi *backbone* pada data uji disajikan secara ringkas dalam Tabel 4.3 yang memuat perbandingan metrik evaluasi utama serta karakteristik kinerja dari setiap model.

Tabel. 4.3 Kinerja YOLOv5 dengan *Backbone* EfficientNet-B0, B1, dan B4

Metrik	EfficientNet-B0	EfficientNet-B1	EfficientNet-B4
Precision	85.3	70.6	77.4
Recall	86.9	88.2	84.7
mAP@0.5	91.9	89.3	90.1
mAP@0.5:0.95	77.7	75.4	74.6
Inference Time	≈ 15.2 ms	≈ 20.0 ms	40.8 ms

Berdasarkan Tabel 4.3, penggunaan backbone EfficientNet pada YOLOv5 menunjukkan bahwa setiap varian memiliki karakteristik performa yang berbeda dalam sistem deteksi nilai mata uang rupiah. EfficientNet-B0 menghasilkan performa terbaik pada sebagian besar metrik evaluasi, dengan precision sebesar 85.3%, recall 86.9%, mAP@0.5 sebesar 91.9%, dan mAP@0.5:0.95 sebesar 77.7%. Selain itu, EfficientNet-B0 juga memiliki waktu inference paling cepat, yaitu sekitar 15,2 ms per gambar. Hasil ini menunjukkan bahwa EfficientNet-B0 mampu memberikan keseimbangan yang baik antara akurasi deteksi dan efisiensi komputasi, sehingga proses deteksi dapat dilakukan secara lebih cepat tanpa mengurangi kualitas deteksi secara signifikan.

EfficientNet-B1 menunjukkan peningkatan kemampuan dalam mendeteksi objek yang ditunjukkan oleh nilai recall tertinggi, yaitu 88.2%. Hal ini menunjukkan bahwa model lebih mampu menemukan objek uang rupiah dan mengurangi jumlah objek yang terlewat. Namun, nilai precision, mAP@0.5, dan mAP@0.5:0.95 lebih rendah dibandingkan EfficientNet-B0, yang mengindikasikan bahwa peningkatan kapasitas ekstraksi fitur belum sepenuhnya menghasilkan

kualitas klasifikasi dan lokalisasi objek yang lebih baik. Kondisi ini dapat disebabkan oleh karakteristik dataset serta kemiripan visual antar nominal uang rupiah yang menyebabkan model menghasilkan deteksi yang kurang konsisten pada beberapa kasus.

Sementara itu, EfficientNet-B4 menghasilkan performa yang tetap kompetitif dengan nilai precision sebesar 77,4%, recall 84,7%, mAP@0.5 sebesar 90,1%, dan mAP@0.5:0.95 sebesar 74,6%. Hasil ini menunjukkan bahwa backbone yang lebih dalam dan kompleks mampu menghasilkan representasi fitur yang kaya sehingga model tetap dapat mengenali detail visual uang rupiah seperti angka nominal, pola ornamen, warna, dan tekstur dengan baik. Namun, peningkatan kompleksitas arsitektur tersebut tidak selalu diikuti oleh peningkatan kinerja dibandingkan varian yang lebih ringan. Selain itu, waktu inference meningkat menjadi 40,8 ms per gambar, yang menunjukkan adanya trade-off antara kompleksitas model dan efisiensi komputasi.

Peningkatan recall menunjukkan bahwa model dengan backbone EfficientNet mampu mendeteksi lebih banyak objek uang yang terdapat dalam citra sehingga jumlah objek yang terlewat (*false negative*) dapat dikurangi. Kemampuan ini berkaitan dengan mekanisme *compound scaling* pada EfficientNet yang meningkatkan kapasitas jaringan secara seimbang melalui penyesuaian tiga dimensi utama, yaitu *depth* (kedalaman jaringan), *width* (jumlah kanal fitur), dan *resolution* (resolusi citra masukan). Dengan meningkatnya ketiga dimensi tersebut secara simultan, model mampu mengekstraksi fitur visual yang lebih representatif

sehingga objek uang dapat dikenali dengan lebih baik pada berbagai kondisi citra. Prinsip *compound scaling* pada EfficientNet dinyatakan melalui persamaan berikut:

$$\text{Depth} : d = \alpha^{\phi} \ , \ \text{Width} : w = \beta^{\phi} \ , \ \text{Resolution} : r = \gamma^{\phi}$$

Meskipun EfficientNet-B4 memiliki arsitektur yang lebih kompleks dibandingkan EfficientNet-B0 dan B1, peningkatan kapasitas model tidak selalu menghasilkan kinerja deteksi yang lebih tinggi. Pada penelitian ini, EfficientNet-B0 justru memperoleh nilai mAP@0.5 dan mAP@0.5:0.95 yang lebih baik. Kondisi tersebut dapat disebabkan oleh ukuran dataset yang relatif terbatas sehingga model dengan jumlah parameter yang lebih besar berpotensi mengalami overfitting. Selain itu, dataset yang digunakan memiliki variasi kondisi citra yang cukup tinggi, seperti pencahayaan yang tidak merata, uang yang terlipat, kusut, lusuh, perubahan orientasi objek, serta latar belakang yang kompleks. Variasi tersebut menyebabkan distribusi fitur menjadi lebih beragam sehingga peningkatan kompleksitas model belum tentu mampu menghasilkan generalisasi yang lebih baik. Faktor lain seperti ketidakseimbangan jumlah data antar kelas serta konfigurasi hyperparameter yang belum sepenuhnya optimal juga dapat memengaruhi penurunan performa pada EfficientNet-B1 dan EfficientNet-B4. Dengan demikian, kompleksitas model yang lebih tinggi tidak selalu memberikan keuntungan yang signifikan ketika dihadapkan pada jumlah data yang terbatas dan kondisi citra yang tidak ideal.

Feature map yang lebih detail membantu YOLOv5 dalam mengenali berbagai karakteristik visual pada uang rupiah, seperti pola angka nominal, tekstur kertas uang, serta perbedaan warna antar pecahan uang. Dengan fitur yang lebih lengkap, model menjadi lebih mudah menemukan objek uang yang sebelumnya

sulit terdeteksi. Akibatnya, jumlah objek yang berhasil dideteksi meningkat sehingga nilai recall juga meningkat. Namun demikian, peningkatan kapasitas jaringan tidak selalu diikuti oleh peningkatan precision. Hal ini karena precision lebih dipengaruhi oleh kemampuan model dalam membedakan antar kelas objek, sedangkan peningkatan *depth* dan *width* pada EfficientNet lebih banyak berkontribusi pada kemampuan model dalam menemukan objek yang ada pada citra. Dengan demikian, peningkatan nilai recall pada EfficientNet-B4 menunjukkan bahwa metode *compound scaling* berhasil meningkatkan kemampuan model dalam menemukan objek uang pada citra, meskipun terjadi sedikit penururunan precision.

Analisis selanjutnya bahwa peningkatan ukuran backbone EfficientNet menyebabkan peningkatan waktu inferensi model. EfficientNet-B0 memiliki waktu inferensi sekitar 15.2 ms, EfficientNet-B1 sekitar 20.0 ms, sedangkan EfficientNet-B4 mencapai sekitar 40.8 ms. Hal ini menunjukkan bahwa semakin besar model yang digunakan, maka semakin lama waktu yang dibutuhkan untuk memproses satu citra. Peningkatan waktu inferensi tersebut berkaitan langsung dengan prinsip *compound scaling* yang meningkatkan kapasitas jaringan melalui tiga dimensi utama dimana pada persamaan tersebut, peningkatan nilai ϕ menyebabkan jaringan menjadi lebih dalam, lebih lebar, dan menggunakan resolusi citra yang lebih tinggi.

Ketika model berubah dari EfficientNet-B0 menjadi EfficientNet-B1 dan kemudian EfficientNet-B4, jumlah layer jaringan meningkat, jumlah kanal fitur bertambah, serta ukuran citra yang diproses oleh model menjadi lebih besar. Peningkatan ketiga dimensi tersebut menyebabkan jumlah operasi komputasi yang

harus dilakukan oleh jaringan juga meningkat. Secara umum, kompleksitas komputasi pada jaringan CNN berbanding lurus dengan kedalaman jaringan, jumlah kanal fitur, dan resolusi citra yang diproses. Oleh karena itu, model dengan kapasitas yang lebih besar seperti EfficientNet-B4 memerlukan waktu pemrosesan yang lebih lama dibandingkan model yang lebih kecil seperti EfficientNet-B0. Pada penelitian ini, peningkatan kompleksitas model tidak selalu diikuti oleh peningkatan kinerja deteksi. Meskipun EfficientNet-B4 memiliki kemampuan representasi fitur yang lebih tinggi, EfficientNet-B0 justru menghasilkan nilai mAP@0.5 dan mAP@0.5:0.95 yang lebih baik. Kondisi ini menunjukkan bahwa model yang lebih kompleks belum tentu memberikan performa terbaik pada dataset dengan jumlah data yang terbatas, kondisi citra yang tidak ideal, serta latar belakang yang kompleks. Hal ini sejalan dengan studi oleh Tan dan Le yang menunjukkan bahwa peningkatan skala EfficientNet dari varian yang lebih kecil ke varian yang lebih besar akan meningkatkan jumlah parameter, kompleksitas komputasi (FLOPs), dan waktu inferensi, namun peningkatan kapasitas model tersebut tidak selalu menghasilkan peningkatan akurasi yang sebanding pada setiap dataset dan tugas deteksi objek [19].

Dengan demikian, pemilihan backbone pada sistem deteksi objek perlu mempertimbangkan keseimbangan antara akurasi deteksi dan efisiensi komputasi. Pertimbangan ini menjadi penting, terutama pada aplikasi yang menuntut proses deteksi secara real-time, di mana peningkatan performa deteksi harus diimbangi dengan waktu inferensi yang tetap efisien.

4.4 Perbandingan Hasil Penelitian dengan Penelitian Terdahulu

Hasil penelitian menunjukkan bahwa penggunaan EfficientNet-B4 sebagai backbone pengganti CSPDarknet pada YOLOv5 mampu meningkatkan performa deteksi objek dibandingkan model baseline. Peningkatan tersebut ditunjukkan oleh kenaikan precision dari 74,6% menjadi 77,4%, recall dari 59,6% menjadi 84,7%, mAP@0.5 dari 69,5% menjadi 90,1%, serta mAP@0.5:0.95 dari 48,0% menjadi 74,6%. Hasil ini menunjukkan bahwa EfficientNet-B4 mampu menghasilkan representasi fitur yang lebih kaya sehingga kemampuan deteksi dan kualitas lokalisasi objek menjadi lebih baik. Peningkatan recall yang signifikan menunjukkan bahwa model mampu mendeteksi lebih banyak objek uang rupiah yang terdapat dalam citra, sedangkan peningkatan mAP@0.5:0.95 menunjukkan bahwa prediksi bounding box yang dihasilkan menjadi lebih akurat meskipun dievaluasi pada threshold IoU yang lebih ketat.

Jika dibandingkan dengan penelitian Wang et al. (2024) yang menggunakan YOLOv5 pada dataset VisDrone2019, penelitian tersebut memperoleh mAP@0.5 sebesar 39,2% dan mAP@0.5:0.95 sebesar 22,3%. Selain itu, penelitian Dontabhaktuni et al. (2024) yang mengevaluasi YOLOv5 dengan variasi *batch size* dan epoch memperoleh mAP@0.5 sebesar 82,1% dan mAP@0.5:0.95 sebesar 50,4%. Meskipun perbandingan secara langsung tidak dapat dilakukan karena perbedaan dataset, karakteristik objek, dan skenario pengujian, hasil tersebut menunjukkan bahwa performa deteksi sangat dipengaruhi oleh kualitas fitur yang dihasilkan backbone serta kompleksitas dataset yang digunakan.

Penelitian Mahasin dan Dewi (2022) yang membandingkan CSPDarkNet53, CSPResNeXt-50, dan EfficientNet-B0 pada YOLOv4 menunjukkan bahwa penggunaan backbone EfficientNet-B0 menghasilkan performa yang lebih baik dibandingkan backbone lainnya. Temuan tersebut mengindikasikan bahwa pemilihan backbone memiliki pengaruh terhadap kualitas ekstraksi fitur dan performa deteksi objek. Hasil penelitian ini memperkuat temuan tersebut, di mana penggunaan EfficientNet-B4 pada YOLOv5 mampu meningkatkan kemampuan deteksi dan kualitas lokalisasi objek dibandingkan YOLOv5 baseline. Peningkatan tersebut ditunjukkan oleh kenaikan nilai mAP@0.5:0.95 dari 48.0% menjadi 74.6% atau terjadi peningkatan sebesar 55.41%, yang menunjukkan bahwa model mampu menghasilkan prediksi *bounding box* yang lebih akurat pada berbagai threshold IoU. Dengan demikian, penggunaan backbone EfficientNet terbukti mampu meningkatkan kualitas representasi fitur yang berkontribusi terhadap peningkatan performa deteksi objek.

Posisi penelitian ini dibandingkan penelitian sebelumnya terletak pada fokus penelitian yang secara khusus mengevaluasi pengaruh penggunaan EfficientNet-B4 terhadap kualitas ekstraksi fitur dan ketepatan lokalisasi objek pada deteksi nilai mata uang rupiah. Berbeda dengan penelitian terdahulu yang umumnya berfokus pada peningkatan akurasi deteksi secara umum atau perbandingan backbone pada dataset lain, penelitian ini mengevaluasi kualitas lokalisasi objek menggunakan metrik mAP@0.5 dan mAP@0.5:0.95 serta menerapkan strategi pelatihan *freeze–unfreeze* untuk mengoptimalkan proses transfer learning. Dengan demikian, kontribusi utama penelitian ini adalah

memberikan bukti empiris bahwa penggunaan EfficientNet-B4 yang dikombinasikan dengan strategi *freeze–unfreeze* mampu meningkatkan kualitas representasi fitur dan presisi bounding box pada sistem deteksi nilai mata uang rupiah berbasis YOLOv5.

BAB 5
PENUTUP

5.1 Kesimpulan

Berdasarkan rumusan masalah dan hasil penelitian yang dilakukan, diperoleh beberapa kesimpulan antara lain:

a. Kinerja model YOLOv5 baseline dalam mendeteksi nilai mata uang rupiah menunjukkan bahwa model mampu melakukan deteksi objek dengan baik pada citra uang rupiah. Model YOLOv5 mampu mengenali berbagai pecahan uang kertas rupiah dengan memanfaatkan proses *ekstraksi fitur* pada *backbone* CSPDarknet yang menghasilkan *Feature map* sebagai dasar proses deteksi objek. Hasil pengujian menunjukkan bahwa model YOLOv5 baseline memiliki kemampuan deteksi yang cukup baik dalam mengidentifikasi objek uang serta menentukan lokasi objek melalui prediksi *bounding box* dan klasifikasi kelas objek.
b. Kinerja model YOLOv5 dengan *backbone* EfficientNet-B4 menunjukkan peningkatan performa deteksi dibandingkan model YOLOv5 *baseline*. Backbone EfficientNet-B4 mampu menghasilkan *Feature map* yang lebih informatif karena menggunakan pendekatan *compound scaling* yang meningkatkan kedalaman jaringan, jumlah channel *feature map*, serta resolusi citra input secara seimbang. Peningkatan kapasitas jaringan ini memungkinkan model mengekstraksi karakteristik visual uang rupiah secara lebih detail, seperti

angka nominal, pola ornamen, gambar tokoh, dan tekstur keamanan, sehingga meningkatkan ketepatan deteksi objek.

c. Penggantian *backbone* YOLOv5 dari CSPDarknet menjadi EfficientNet-B4 memberikan pengaruh positif terhadap kinerja sistem deteksi nilai mata uang rupiah. *Backbone* EfficientNet-B4 mampu menghasilkan representasi fitur yang lebih baik sehingga membantu model dalam mengenali objek dengan lebih akurat. Hal ini tercermin dari peningkatan nilai metrik evaluasi seperti recall, dan mean Average Precision (mAP). Dengan demikian, penggunaan *backbone* EfficientNet-B4 dapat meningkatkan kemampuan sistem deteksi objek berbasis YOLOv5 dalam mengenali dan membedakan pecahan uang rupiah.

d. Penggunaan EfficientNet-B4 berhasil meningkatkan performa deteksi, namun menyebabkan waktu inferensi menjadi lebih tinggi sehingga kurang optimal untuk aplikasi yang membutuhkan respons *real-time* dengan latensi rendah.

e. Perbandingan kinerja EfficientNet-B4 dengan EfficientNet-B0 dan EfficientNet-B1 menunjukkan bahwa model dengan backbone EfficientNet-B4 memiliki performa deteksi yang lebih baik dibandingkan kedua varian EfficientNet lainnya. Hal ini menunjukkan bahwa peningkatan kapasitas jaringan melalui metode *compound scaling* berkontribusi terhadap peningkatan kualitas *Feature map* yang dihasilkan oleh *backbone* CNN. *Feature map* yang lebih kaya informasi memungkinkan model mendeteksi objek dengan lebih akurat serta membedakan kelas objek yang memiliki kemiripan visual.

f. Secara teoritis, penelitian ini memberikan kontribusi dalam menunjukkan bahwa penerapan metode *compound scaling* pada backbone EfficientNet dapat

meningkatkan kualitas representasi fitur pada jaringan CNN yang digunakan dalam sistem deteksi objek berbasis YOLOv5. Temuan ini memperkuat pemahaman bahwa pemilihan arsitektur *backbone* memiliki peran penting dalam menentukan performa model deteksi objek.

g. Secara praktis, penelitian ini memberikan kontribusi dalam pengembangan sistem deteksi nilai mata uang rupiah berbasis *deep learning* yang dapat digunakan sebagai dasar pengembangan aplikasi pengenalan uang secara otomatis, misalnya pada sistem pendukung bagi penyandang tunanetra, sistem kasir otomatis, atau perangkat identifikasi nilai mata uang berbasis komputer vision.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa hal yang dapat dikembangkan pada penelitian selanjutnya:

a. Penelitian berikutnya dapat memperluas cakupan deteksi tidak hanya pada nominal uang kertas, tetapi juga pada berbagai fitur keamanan pada uang rupiah, seperti watermark, microtext, pola ornamen, maupun elemen visual lain yang digunakan sebagai tanda keaslian uang.
b. Pada penelitian selanjutnya, eksplorasi pengembangan model tidak hanya difokuskan pada penggantian backbone, tetapi juga dapat diarahkan pada modifikasi bagian neck maupun detection head pada arsitektur YOLO.
c. Sistem deteksi yang dikembangkan dalam penelitian ini masih berfokus pada proses deteksi objek pada citra statis. Penelitian selanjutnya dapat mengembangkan sistem ini ke dalam bentuk aplikasi berbasis perangkat

bergerak atau sistem *real-time*, sehingga model dapat digunakan secara langsung dalam berbagai kebutuhan praktis, seperti alat bantu pengenalan uang bagi penyandang tunanetra atau sistem identifikasi mata uang pada perangkat otomatis.

d. Pada penelitian selanjutnya, pendekatan yang digunakan dalam penelitian ini juga berpotensi dikembangkan untuk aplikasi lain yang membutuhkan tingkat ketelitian lokalisasi objek yang tinggi, misalnya pada deteksi tumor otak pada citra MRI. Hal ini karena penggunaan backbone EfficientNet-B4 pada penelitian ini menunjukkan kestabilan proses pelatihan serta peningkatan akurasi deteksi. Dengan kemampuan ekstraksi fitur yang baik, arsitektur tersebut berpotensi diterapkan pada aplikasi medis yang lebih menekankan presisi deteksi dibandingkan kecepatan pemrosesan, tentu dengan dukungan dataset medis dan sumber daya komputasi yang memadai.

DAFTAR PUSTAKA

[1] Md. Jubayar Alam Rafi, M. Rony & Nazia Majadi, ''NSTU-BDTAKA: An open dataset for Bangladeshi paper currency detection and recognition,'' *Elsevier*. 2024.

[2] Joseph Redmon, Santosh Divvala, Ross Girshick & Ali Farhadi, ''You Only Look Once: Unified, Real-Time Object Detection,'' *Available: http://arxiv.org/abs/1506.02640*. 2016.

[3] Ardanu D Nugroho & Wiga M Baihaqi, ''Improved YOLOv5 with *Backbone* Replacement to MobileNet V3s for School Attribute Detection,'' *SinkrOn*. 2023.

[4] Tausif Diwan, G. Anirudh & Jitendra V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Springer*, 2022.

[5] Leo Thomas Ramos & Angel D. Sappa, "A Decade of You Only Look Once (YOLO) for Object Detection," *arXiv:2504.18586v1*, 2025.

[6] Xinzhu Ma et al, "Delving into Localization Errors for Monocular 3D Object Detection," *arXiv:2103.16237v1*, 2021.

[7] S Divya Meena, Veeramachaneni Gayathri siva sameeraja, Nagineni Sai Lasya, Meda Sathvika, Veluru Harshitha & J Sheela, ''Hybrid Neural Network Architecture For Multi-Label Object Recognition Using Feature Fusion,'' *Elsevier*. 2023

[8] Mengshu JIAO, Jianbiao HE & Baojiang ZHANG, ''Folding Paper Currency Recognition And Research Based On Convolution Neural Network,'' *IEEE*. 2018.

[9] A. Nasayreh, Ameera S. Jaradat & Hasan Gharaibeh et al, ''Jordanian Banknote Data Recognition: A Cnn-Based Approach With Attention Mechanism,'' *Elsevier*. 2024.

[10] Bo Xu, Bin Gao & Yunhu Li, ''Improved Small Object Detection Algorithm Based On YOLOv5,'' *IEEE Computer Society*. 2024.

[11] Siti Nurmaini and Dr. Ahmad Rizki Pratama, ''Improving Object Detection Efficiency: Yolo V4 And *Backbone* Selection,'' *AIJMSE*. 2023.

[12] Marsa Mahasin and Irma Amelia Dewi, ''Comparison of CSPDarkNet53, CSPResNeXt-50, and EfficientNet-B0 *Backbone*s on YOLO V4 as Object Detector,'' *IJESTY*. 2022.

[13] Xuecheng Li, Xiaobin Li, Biao Han, Shang Wang and Kairun Chen, ''Application of EfficientNet and YOLOv5 Model in Submarine Pipeline Inspection and a New Decision-Making System,'' *MDPI*. 2023.

[14] Andrzej D. Dobrzycki, Ana M. Bernardos and José R. Casar, ''An Analysis of Layer-Freezing Strategies for Enhanced Transfer Learning in YOLO

Architectures,'' *MDPI*. 2025.

[15] A. Murat and Mustafa S. Kiran, ''A Comperhensive Review on Yolo Version for Object Detector,'' *Elsevier*. 2025.

[16] Haiying Liu & Fengqian Sun et al, ''Lightweight small object detection algorithm based on improved feature fusion mode,'' *https://doi.org/10.3390/s22155817*. 2022.

[17] Octavian Ery Pamungkas, Puspa Rahmawati, & Dhany Maulana Supriadi et al, ''Classification Of Rupiah To Help Blind With The *Convolutional Neural Network* Method,'' *Shinta2*. 2022.

[18] Alexandr Pak, Atabay Ziyaden & Kuanysh Tukeshev et al, ''Comparative Analysis Of Deep Learning Methods Of Detection Of Diabetic Retinopathy,'' *https://doi.org/10.1080/23311916.2020.1805144*. 2020.

[19] M. Tan and Q. V. Le, ''Efficientnet: Rethinking Model Scaling For *Convolutional Neural Networks*,'' *Available: http://arxiv.org/abs/1905.11946*. 2019.

[20] Kyriakos D. Apostolidis and George A. Papakostas, ''Delving into YOLO Object Detection Models: Insights into Adversarial Robustness,'' *MDPI*. 2025.

[21] Kevin Maulana Azhar, Imam Santoso & Yosua Alvin Adi Soetrisno, ''Implementasi Deep Learning Menggunakan Metode *Convolutional Neural Network* dan Algoritma Yolo Dalam Sistem Pendeteksi Uang Kertas Rupiah Bagi Penyandang Low Vision,'' *Transient, Vol. 10, No. 3*. 2021.

[22] Rakesh Chandra Joshi, Saumya Yadaf & Malay K Dutta, ''Yolov3 Based Currency Detection And Recognition System For Visually Impaired Person,'' *IEEE*. 2020.

[23] Weihai Sun, Yane Li & Hailin Feng et al, ''Lightweight And Accurate Aphid Detection Model Based On An Improved Deep-Learning Network,'' *Elsevier*. 2024.

[24] Peng Wang, Shixin Li, Weixiao Wang, Kaiwen Tang and Fankai Chen, ''Metal Defect Detection Models Fused EfficientNet and Involution,'' *https://doi.org/10.1155/js/6074853*. 2024.

[25] Mingxing Tan and Quoc V. Le, ''EfficientNetV2: Smaller Models and Faster Training,'' *Available: http://arxiv.org/abs/2104.00298v3*. 2021.