

BAB III METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian ini adalah model deteksi objek berbasis *Mask R-CNN* yang dimodifikasi dengan mengganti *backbone* default (*ResNet-50*) menjadi *MobileNetV3 Small*. Fokus utama penelitian adalah untuk meningkatkan efisiensi model deteksi objek dalam konteks gudang distribusi minuman di Gudang Distribusi Minuman Air Toko Anugerah, yang berlokasi di Kampung Kedung Lotong, Desa Bantarjaya, Kabupaten Bekasi. Objek yang akan dideteksi dalam penelitian ini adalah kardus produk minuman dari empat merek, yaitu Sky, Fatqua, Levios, dan Elvios yang dapat dilihat pada gambar 3.1.



Kardus Sky



Kardus Fatqua



Kardus Levios



Kardus Elvios

Gambar 3. 1 Objek Penelitian

Sumber: Dokumentasi Peneliti (2025)

Kardus-kardus ini diambil dari berbagai sudut dan kondisi pencahayaan untuk memastikan variasi dalam *dataset*. Dengan mengganti *backbone Mask R-*

CNN menjadi *MobileNetV3 Small*, penelitian ini bertujuan untuk mengoptimalkan kecepatan inferensi dan efisiensi penggunaan memori tanpa mengorbankan akurasi deteksi.

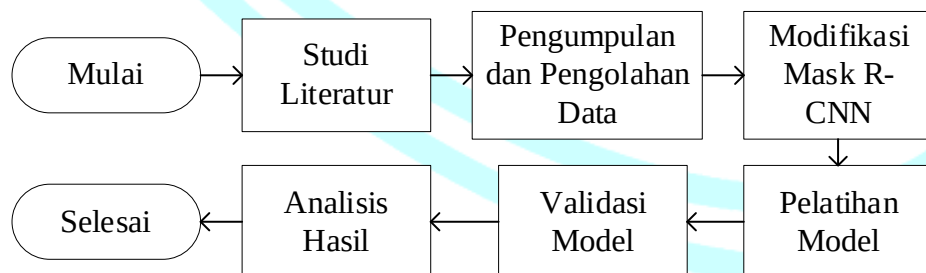
Penelitian ini dilaksanakan di Laboratorium Riset Universitas Buana Perjuangan Karawang selama lima bulan, mulai November 2024 hingga Maret 2025. Rincian tahapan penelitian beserta alokasi waktunya disajikan secara terperinci pada tabel yang tersedia.

Tabel 3. 1 Waktu Penelitian

No	Kegiatan	2024												2025											
		11				12				1				2				3							
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4				
1	Studi Literatur																								
2	Pengumpulan Data & Pengolahan Data																								
3	Modifikasi <i>Mask R-CNN</i>																								
4	Pelatihan																								
5	Validasi																								
6	Analisis																								

3.2 Prosedur Penelitian

Prosedur penelitian ini disusun secara sistematis dalam beberapa tahapan utama untuk mencapai tujuan penelitian secara terarah. Secara umum, alur prosedur dapat dilihat pada Gambar 3.2.



Gambar 3. 2 Prosedur Penelitian

Sumber: Dokumentasi Peneliti (2025)

Adapun tahapan-tahapan penelitian dijelaskan sebagai berikut:

3.2.1 Studi Literatur

Studi literatur atau studi pustaka terhadap jurnal-jurnal yang relevan, khususnya pada jurnal yang membahas tentang algoritma *Mask R-CNN*, *MobileNetV3*, teknik anotasi pada gambar dan *Framework computer vision*. Kajian

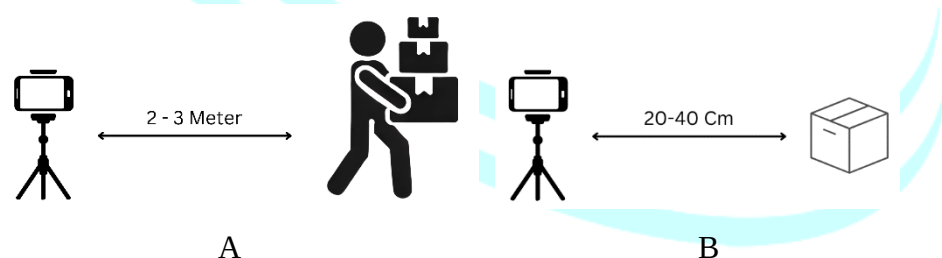
ini melibatkan penelaahan terhadap jurnal-jurnal ilmiah yang relevan guna memperoleh wawasan mendalam mengenai performa *Mask R-CNN* dalam segmentasi objek, efisiensi *MobileNetV3* dalam aplikasi berbasis perangkat seluler, serta teknik-teknik terbaik dalam melakukan anotasi data gambar untuk meningkatkan akurasi model. Selain itu, studi ini juga mencakup analisis terhadap berbagai *framework computer vision* seperti *Detectron2* dan *TensorFlow* yang sering digunakan untuk mempermudah implementasi algoritma canggih. Dengan memadukan temuan dari jurnal-jurnal tersebut, studi pustaka ini diharapkan dapat memberikan landasan teori yang kuat untuk pengembangan metode dan aplikasi berbasis *computer vision* yang lebih optima

3.2.2 Pengumpulan dan Pengolahan Data

Pengumpulan dan pengolahan data merupakan langkah awal yang krusial dalam penelitian ini untuk memastikan kualitas *dataset* yang digunakan dalam pelatihan model. Proses ini terbagi menjadi tiga sub-tahapan utama:

3.2.2.1 Pengambilan Foto Kardus

Pengambilan data dilakukan di Gudang Distribusi Minuman Air Toko Anugerah yang terletak di Kampung Kedung Lotong, Desa Bantarjaya, Kabupaten Bekasi. Foto kardus diambil dari berbagai sudut pandang seperti depan, samping, dan atas untuk menangkap berbagai aspek objek. Kondisi pencahayaan yang bervariasi juga diatur untuk mensimulasikan lingkungan nyata.



Gambar 3. 3 Pengambilan Gambar Kardus

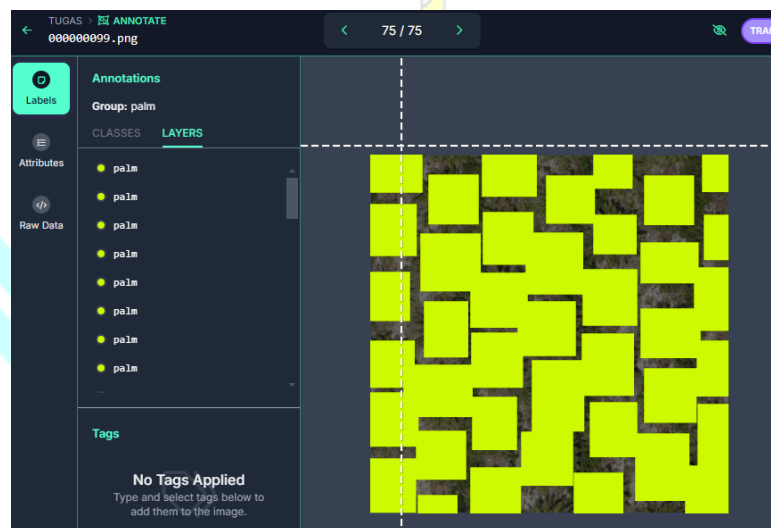
Sumber: Dokumentasi Peneliti (2025)

Foto diambil menggunakan kamera smartphone dengan spesifikasi 48 megapiksel. Jarak pengambilan gambar adalah 20-40 cm untuk kardus satuan yang dapat dilihat pada Gambar 3.2 gambar A dan 2-3 meter untuk kardus yang dibawa oleh karyawan pada gambar B. Kondisi pencahayaan diatur dengan lampu ruangan 15 watt untuk pengambilan kardus dalam ruangan dan pencahayaan alami sekitar

pukul 8 pagi untuk kardus yang dibawa karyawan. Total pengumpulan data yaitu 615 foto dari seluruh kategori yaitu Sky, Fatqua, Levios, dan Elvios, memastikan variasi yang cukup dengan mengambil foto dari setiap sisi produk secara merata.

3.2.2.2 Anotasi Data di Roboflow

Setelah pengambilan foto, langkah selanjutnya adalah melakukan anotasi pada *dataset* untuk menandai objek yang akan dideteksi oleh model. Proses anotasi dilakukan dengan memberi label pada setiap gambar menggunakan alat di situs web *Roboflow* (Karyanto et al., 2024). *Dataset* yang telah diambil diunggah ke platform *Roboflow*, di mana anotasi dilakukan secara manual oleh peneliti sendiri menggunakan fitur segmentasi. Kategori anotasi ditentukan berdasarkan produk yaitu "Sky", "Fatqua", "Levios", dan "Elvios". Proses anotasi memastikan setiap objek kardus terlabeli dengan akurat sesuai produk masing-masing.



Gambar 3. 4 Anotasi pada *Roboflow*

Sumber: Dokumentasi Peneliti (2025)

Pada *Roboflow* di tahap akhir *dataset* dibagi menjadi 2 yaitu untuk pelatihan dan validasi dengan rasio 88% untuk pelatihan dan 12% untuk validasi. Serta menambahkan *augmentasi* yang tersedia pada *Roboflow* yaitu *rotation*, *brightness* dan *blur*. Setiap *instance* dalam *dataset* akan diperbanyak sebanyak tiga kali melalui *augmentasi* tersebut, sehingga total jumlah data yang dihasilkan meningkat secara signifikan.

3.2.2.3 Penyimpanan Dataset di Google Drive

Setelah proses anotasi selesai, *dataset* diekspor dalam format yang kompatibel dengan *Detectron2* untuk memudahkan integrasi dalam *pipeline* pelatihan. *Dataset* yang telah dianotasi diekspor dalam format *COCO JSON* dan disimpan sepenuhnya di Google Drive.

3.2.3 Modifikasi Mask R-CNN dengan MobileNetV3

Modifikasi *Mask R-CNN* bertujuan untuk mengganti *backbone* default *Mask R-CNN* dengan *MobileNetV3 Small*, sehingga meningkatkan efisiensi model tanpa mengurangi akurasi deteksi. Modifikasi ini dilakukan melalui dua sub-tahapan utama:

3.2.3.1 Integrasi Backbone MobileNetV3 Small

MobileNetV3 Small dipilih sebagai *backbone* alternatif karena arsitekturnya yang ringan dan efisien dalam penggunaan sumber daya. *MobileNetV3 Small* diimplementasikan sebagai *backbone* dalam *framework Detectron2* dengan melakukan perubahan pada file konfigurasi dan arsitektur model. Langkah ini melibatkan penggantian *layer backbone* default *Mask R-CNN* dengan *layer MobileNetV3 Small*. Selain itu, penyesuaian dilakukan untuk memastikan kompatibilitas arsitektur *MobileNetV3 Small* dengan head *RPN (Region proposal network)* dan *ROI (Region of interest)* yang digunakan dalam *Mask R-CNN*. Proses kompatibilitas dilakukan dengan menyesuaikan parameter dan komponen yang diperlukan dalam *framework Detectron2*.

3.2.3.2 Penyesuaian Hyperparameter

Penggantian *backbone* dapat mempengaruhi dinamika pelatihan model, sehingga perlu dilakukan penyesuaian pada *hyperparameter* untuk mengakomodasi *backbone* baru. *Hyperparameter* yang disesuaikan meliputi:

1. *Learning rate*: Ditentukan sebesar 0.002 dengan strategi *step decay* pada *iterasi* ke-12.000 dan 16.000, serta *warm-up* selama 1.000 *iterasi* awal untuk menjaga stabilitas pada tahap awal pelatihan.
2. *Batch size*: Ditetapkan sebanyak 2 gambar per *iterasi*, menyesuaikan kapasitas memori dan efisiensi pelatihan pada *backbone MobileNetV3 Small*.

3. Jumlah *Iterasi* Maksimum: Diperpanjang hingga 20.000 *iterasi* untuk memberikan waktu pembelajaran yang cukup terhadap arsitektur *backbone* yang lebih ringan.
4. Proposal *ROI (Region of interest)*: Ditetapkan sebanyak 128 proposal per gambar, sesuai default umum *instance segmentation* dan seimbang antara akurasi serta kecepatan.
5. Ukuran Gambar Input: Disesuaikan dalam rentang 800 hingga 1.333 piksel untuk proses pelatihan dan pengujian, menjaga kompatibilitas dengan ukuran *anchor FPN*.
6. *Backbone* Custom: Menggunakan *MobileNetV3 Small* yang dimodifikasi dengan *Feature Pyramid Network (FPN)* dan *output channels* sebesar 256, untuk menyesuaikan dengan struktur *Mask R-CNN*.
7. Data Loader: Menggunakan 2 *worker* untuk mempercepat proses pemuatan data selama pelatihan.
8. Perangkat Pelatihan: Pelatihan dilakukan secara adaptif, yaitu dengan GPU jika tersedia, atau CPU sebagai alternatif.

Penyesuaian *hyperparameter* dilakukan berdasarkan kondisi yang ada dan hasil dari proses pelatihan yang gagal. *Hyperparameter* disesuaikan secara iteratif untuk mencapai konfigurasi optimal yang mendukung performa model.

3.2.4 Pelatihan Model

Pelatihan model merupakan tahap di mana model yang telah dimodifikasi dilatih menggunakan *dataset* yang telah disiapkan. Proses ini dibagi menjadi dua sub-tahapan utama:

3.2.4.1 Dataset yang digunakan

Model dilatih menggunakan *dataset* kardus untuk meningkatkan kemampuan deteksinya. *Dataset* ini mencakup berbagai kategori produk, yaitu Sky, Fatqua, Levios, dan Elvios, yang telah dianotasi, di *augmentasi* serta dibagi menjadi *dataset* pelatihan dan validasi menggunakan *Roboflow* untuk mendukung proses pelatihan dan validasi model.

3.2.4.2 Proses Pelatihan

Pelatihan dilakukan menggunakan *framework Detectron2* dengan konfigurasi pada subbab 3.5.2 Penyesuaian *hyperparameter*. *Optimizer* yang digunakan adalah *SGD (Stochastic Gradient Descent)* utama untuk mempercepat konvergensi serta Pelatihan dilakukan menggunakan *Google Colab* dengan runtime tipe T4 (15 GB GPU).

Fokus utama pelatihan adalah pada model *Mask R-CNN* yang telah dimodifikasi dengan *backbone MobileNetV3 Small*. Namun untuk keperluan perbandingan performa, dilakukan pula pelatihan terhadap model *Mask R-CNN* dengan *backbone* default (*ResNet-50*) menggunakan konfigurasi pelatihan yang serupa. Hal ini bertujuan untuk mengevaluasi sejauh mana modifikasi *backbone* dapat meningkatkan efisiensi dan efektivitas dalam mendeteksi objek.

3.2.5 Validasi Model

Validasi model adalah proses penting untuk memastikan bahwa model yang dilatih dapat bekerja dengan baik pada data yang belum pernah dilihat sebelumnya. Tahapan ini terdiri dari dua sub-tahapan utama:

3.2.5.1 Metode Validasi

Validasi dilakukan dengan menggunakan subset dari *dataset* kardus untuk mengukur performa kedua model. Proses validasi mencakup:

1. *Confusion Matrix* digunakan untuk menggambarkan hasil klasifikasi empat kemungkinan keluaran yaitu *true positive* (jumlah data positif benar), *true negative* (jumlah data negative benar), *false positive* (jumlah data negatif yang diklasifikasi benar) dan *false negative* (jumlah data negative yang diklasifikasi salah) (Uly et al., 2023).
2. Metrik evaluasi adalah alat atau metode yang digunakan untuk menilai kinerja suatu model dalam menyelesaikan tugas tertentu, Menurut Ardiansyah et al., (2024) beberapa validasi yang digunakan untuk menilai kinerja model meliputi:
 - a. *Presisi* pada rumus (1) mengacu pada tingkat akurasi hasil yang diperoleh suatu model (Nardilasari et al., 2023).

$$presisi = \frac{true\ positive}{true\ positive + false\ positive} \quad (1)$$

- b. *Recall* pada rumus (2) menunjukkan seberapa lengkap data relevan yang berhasil diidentifikasi oleh sebuah model (Nardilasari et al., 2023).

$$recall = \frac{true\ positive}{true\ positive + false\ negative} \quad (2)$$

- c. *Intersection over Union (IoU)* adalah *metric* yang digunakan untuk mengevaluasi akurasi *bounding box* yang diprediksi model untuk mendeteksi objek (Mansyur & Pratiwi, 2025).

3. Kriteria Keberhasilan: *Threshold* yang digunakan untuk menentukan keberhasilan model adalah 0,5.

3.2.5.2 Experimen Validasi

Model diuji pada gambar yang belum pernah dilihat selama pelatihan untuk mengukur kemampuan generalisasi. Model diaplikasikan pada gambar baru yang diambil dari lingkungan berbeda, memastikan bahwa model tidak hanya menghafal data pelatihan tetapi juga dapat mendeteksi objek dalam kondisi nyata. Sebagai bagian dari eksperimen validasi dilakukan pula perbandingan hasil inferensi antara dua model: *Mask R-CNN* default dengan *backbone ResNet-50* dan model modifikasi dengan *backbone MobileNetV3 Small*. Hasil inferensi ini diaplikasikan pada gambar uji yang sama sehingga perbedaan performa dapat dianalisis secara visual dan kuantitatif.

Validasi ini memastikan bahwa model tidak hanya berkinerja baik pada data pelatihan tetapi juga dapat diandalkan dalam situasi nyata dengan data yang bervariasi.

3.2.6 Analisis

Setelah model divalidasi, tahap selanjutnya adalah menganalisis hasil untuk menentukan efektivitas penggantian *backbone Mask R-CNN* dengan *MobileNetV3 Small*.

Performa *backbone MobileNetV3 Small* dibandingkan dengan *backbone* default *Mask R-CNN* untuk menilai peningkatan atau penurunan kinerja. Implementasinya meliputi:

1. Kecepatan Inferensi (*FPS*): Mengukur berapa banyak frame per detik yang dapat diproses oleh model, memberikan indikasi tentang seberapa cepat model dapat digunakan dalam aplikasi waktu nyata.
2. Akurasi: Membandingkan nilai *Confusion Matrix* dan Metrik Evaluasi pada model untuk menilai akurasi deteksi objek.
3. Ukuran Model: Menilai ukuran file model untuk menentukan efisiensi penggunaan memori, yang penting untuk aplikasi yang berjalan pada perangkat dengan sumber daya terbatas.

