

**PENGEMBANGAN SISTEM OCR BERBASIS CRAFT DAN
CRNN UNTUK PENGENALAN TEKS DAFTAR
BAHAN PADA LABEL MAKANAN**



Disusun Oleh:

Nama : Rakeen Aria Alireza
NIM : 22523252

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
2026**

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**PENGEMBANGAN SISTEM OCR BERBASIS CRAFT DAN
CRNN UNTUK PENGENALAN TEKS DAFTAR
BAHAN PADA LABEL MAKANAN**



Yogyakarta, 27 April 2026

Pembimbing,


(Arrie Kurniawati, S.Si., M.Kom.)

HALAMAN PENGESAHAN DOSEN PENGUJI

PENGEMBANGAN SISTEM OCR BERBASIS CRAFT DAN CRNN UNTUK PENGENALAN TEKS DAFTAR BAHAN PADA LABEL MAKANAN

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 27 April 2026

Tim Penguji

Arrie Kurniawardhani, S.Si., M.Kom.

Anggota 1

Dr. Hendrik, S.T., M.Eng.

Anggota 2

Moh. Idris, S.Kom., M.Kom.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Rakeen Aria Alireza

NIM : 22523252

Tugas akhir dengan judul:

PENGEMBANGAN SISTEM OCR BERBASIS CRAFT DAN CRNN UNTUK PENGENALAN TEKS DAFTAR BAHAN PADA LABEL MAKANAN

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 27 April 2026



(Rakeen Aria Alireza)

HALAMAN PERSEMBAHAN

Alhamdulillah rabbil 'alamin, puji dan syukur penulis panjatkan ke hadirat Allah Swt. atas segala nikmat, rahmat, dan kemudahan-Nya sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik. Karya ini penulis persembahkan kepada kedua orang tua tercinta serta seluruh keluarga yang selalu mendukung dan mendoakan penulis. Terima kasih atas segala dukungan, doa yang tidak pernah putus, serta nasihat yang senantiasa menguatkan penulis selama proses penyusunan tugas akhir. Semoga hasil penelitian ini dapat memberikan manfaat bagi pengembangan sistem OCR untuk pembacaan daftar bahan pada label makanan kemasan.

HALAMAN MOTO

“I dream my painting and I paint my dream.”

(Vincent Willem van Gogh)

KATA PENGANTAR

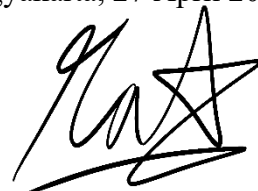
Assalamualaikum Warahmatullahi Wabarakatuh.

Alhamdulillahirabbil'amin, segala puji bagi Allah Swt. atas rahmat, karunia, dan kemudahan-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir ini. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad saw., keluarga, sahabat, dan umatnya hingga akhir zaman.

Laporan tugas akhir ini disusun sebagai salah satu syarat menyelesaikan pendidikan Program Sarjana (S-1) Program Studi Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia. Penelitian ini bertujuan mengembangkan sistem OCR berbasis CRAFT dan CRNN-TRBA untuk pembacaan teks daftar bahan pada label makanan kemasan. Dalam proses penyusunan tugas akhir ini, penulis menghadapi beberapa kendala dalam penelitian, penulisan, dan kewajiban akademik lainnya. Namun, berkat dukungan berbagai pihak, laporan ini dapat diselesaikan dengan baik. Oleh karena itu, penulis hendak menyampaikan terima kasih kepada:

1. Kedua orang tua dan keluarga penulis yang selalu memberi dukungan, doa, serta motivasi.
2. Prof. Fathul Wahid, S.T., M.Sc., Ph.D., selaku Rektor Universitas Islam Indonesia.
3. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D., selaku Ketua Program Studi Informatika Program Sarjana.
4. Ibu Arrie Kurniawardhani, S.Si., M.Kom., selaku dosen pembimbing yang memberi arahan, bimbingan, dan masukan selama penelitian.
5. Bapak dan Ibu dosen Program Studi Informatika yang telah memberikan ilmu dan pandangan luas selama perkuliahan di Informatika UII.
6. Sahabat seperjuangan di lingkungan kampus, yang menjadi rekan bertumbuh melalui diskusi, saling bantu, dan dukungan nyata hingga penelitian ini dapat diselesaikan.
7. Teman-teman masa sekolah penulis, yang tetap hadir sebagai ruang pulang untuk berbagi cerita dan penguat langkah selama proses penyusunan tugas akhir ini.

Yogyakarta, 27 April 2026



(Rakeen Aria Alireza)

SARI

Daftar bahan pada label makanan kemasan umumnya masih tersedia sebagai teks cetak pada citra kemasan, sehingga belum langsung dapat digunakan sebagai data digital yang terstruktur. Proses ekstraksi teks pada label makanan memiliki tantangan karena ukuran huruf kecil, susunan teks padat, variasi font, kualitas citra yang tidak seragam, serta kemungkinan distorsi perspektif dan permukaan kemasan yang melengkung. Penelitian ini bertujuan merancang dan mengevaluasi sistem *Optical Character Recognition* (OCR) untuk mengekstraksi teks daftar bahan dari citra label makanan secara akurat menggunakan kombinasi deteksi teks dan pengenalan teks berbasis pembelajaran mendalam.

Metode penelitian meliputi pengumpulan data, *preprocessing*, pembuatan sistem OCR, *fine-tuning* model pengenalan teks, evaluasi, simulasi eksternal, dan pembuatan antarmuka demonstrasi. Data utama menggunakan dataset *Ingredients Image from Food Label* sebanyak 182 citra yang diekstraksi menjadi 6812 sub-citra kata, kemudian dikurasi menjadi 6081 data valid. Data dibagi ke subset latih, validasi, dan uji dengan rasio 80:10:10, lalu dikonversi ke format *Lightning Memory-Mapped Database* (LMDB). Sistem OCR dibangun menggunakan *Character Region Awareness for Text Detection* (CRAFT) untuk deteksi teks dan *Convolutional Recurrent Neural Network* (CRNN) berarsitektur TPS-ResNet-BiLSTM-Attention (TRBA) untuk pengenalan teks. *Fine-tuning* dilakukan pada model CRNN pra-latih dengan membekukan modul TPS dan ResNet, sedangkan modul BiLSTM dan Attention tetap dilatih. Evaluasi dilakukan menggunakan *Word Accuracy*, *Normalized Edit Distance* (NED), *Character Error Rate* (CER), dan *Word Error Rate* (WER).

Hasil penelitian menunjukkan bahwa *fine-tuning* meningkatkan performa model pada data uji. Model pra-latih menghasilkan *Word Accuracy* 90,381%, NED 0,946, CER 0,066, dan WER 0,096, sedangkan model hasil *fine-tuning* terbaik menghasilkan *Word Accuracy* 94,693%, NED 0,987, CER 0,015, dan WER 0,053. Perbandingan dengan Tesseract menunjukkan bahwa sistem CRAFT dan CRNN-TRBA lebih sesuai untuk data uji penelitian ini karena Tesseract memperoleh *Word Accuracy* 70,813%, NED 0,894, CER 0,089, dan WER 0,300. Pada simulasi lima citra label makanan eksternal, sistem menghasilkan *Word Accuracy* 92,248%, NED 0,988, CER 0,009, dan WER 0,078, sedangkan pada tiga citra label obat menghasilkan *Word Accuracy* 85,382%, NED 0,965, CER 0,034, dan WER 0,146. Dengan demikian, sistem OCR yang dikembangkan mampu menjadi fondasi digitalisasi teks daftar bahan pada label makanan, meskipun masih perlu peningkatan pada variasi data, pascaproses kesalahan karakter,

pengujian eksternal yang lebih luas, dan penyempurnaan antarmuka pada penelitian berikutnya.

Kata kunci: OCR, CRAFT, CRNN-TRBA, *fine-tuning*, label bahan makanan.

GLOSARIUM

Adam	Algoritma optimasi adaptif berbasis gradien untuk memperbarui bobot model saat pelatihan.
<i>Batch size</i>	Jumlah sampel yang diproses dalam satu langkah pembaruan parameter.
<i>Checkpoint</i>	Berkas <i>snapshot</i> bobot model pada iterasi atau kinerja tertentu selama pelatihan.
<i>Dataset</i>	Kumpulan data citra beserta label yang digunakan untuk data latih, validasi, dan uji.
<i>Fine-tuning</i>	Proses penyesuaian model pra-latih menggunakan data target agar kinerjanya sesuai domain penelitian.
<i>Ground truth</i>	Label teks acuan yang dianggap benar untuk pelatihan dan evaluasi.
<i>Inference</i>	Proses menjalankan model terlatih pada data baru untuk menghasilkan prediksi.
<i>Input channel</i>	Jumlah kanal warna pada citra masukan, misalnya 1 kanal (<i>grayscale</i>) atau 3 kanal (RGB).
<i>Iteration</i>	Satu siklus pelatihan yang mencakup <i>forward pass</i> , perhitungan <i>loss</i> , dan pembaruan bobot.
<i>Learning rate</i>	Besaran langkah pembaruan bobot pada setiap iterasi optimasi.
LMDB	Format basis data yang efisien untuk penyimpanan dan pemuatan data OCR.
<i>Preprocessing</i>	Tahap penyiapan data sebelum pelatihan atau inferensi, seperti normalisasi citra, kurasi label, dan konversi format.
Streamlit	Kerangka kerja Python untuk membangun antarmuka demonstrasi sistem OCR.
Sub-citra	Potongan citra dari citra utama yang biasanya memuat satu kata atau satu area teks.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING	ii
HALAMAN PENGESAHAN DOSEN PENGUJI	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....	iv
HALAMAN PERSEMBAHAN.....	v
HALAMAN MOTO.....	vi
KATA PENGANTAR.....	vii
SARI	viii
GLOSARIUM	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Penelitian	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II DAFTAR PUSTAKA.....	6
2.1 Dasar Teori.....	6
2.1.1 <i>Optical Character Recognition (OCR)</i>	6
2.1.2 <i>Character Region Awareness for Text detection (CRAFT)</i>	6
2.1.3 <i>Convolution Recurrent Neural Network (CRNN)</i> dengan Kerangka TPS-ResNet-BiLSTM-Attention (TRBA)	7
2.1.4 <i>Fine-tuning</i>	11
2.1.5 Metrik Evaluasi	11
2.1.6 <i>Lightning Memory-Mapped Database (LMDB)</i>	13
2.1.7 Streamlit	14
2.2 Penelitian Sebelumnya	14
2.3 Pemilihan Metode	15
BAB III METODOLOGI PENELITIAN.....	18
3.1 Langkah-langkah Penelitian.....	18
3.1.1 Pengumpulan Data	19
3.1.2 <i>Preprocessing</i>	20
3.1.3 Pembuatan Sistem OCR.....	20
3.1.4 <i>Fine-tuning</i> Model Pengenalan Teks	22
3.1.5 Evaluasi	24
3.1.6 Pembuatan Antarmuka	24
3.1.7 Simulasi Sistem OCR.....	25
BAB IV HASIL DAN PEMBAHASAN.....	26
4.1 Pengumpulan Data	26
4.1.1 Data Utama Penelitian	26
4.1.2 Data Eksternal Simulasi	27
4.2 <i>Preprocessing</i>	28
4.3 Pembuatan Sistem OCR.....	29
4.3.1 Implementasi Deteksi Teks	29

4.3.2	Implementasi Pengenalan Teks.....	31
4.3.3	Rekonstruksi Teks Label.....	31
4.4	<i>Fine-tuning</i> Model Pengenalan Teks	35
4.4.1	Hasil Pelatihan <i>Fine-tuning</i>	35
4.4.2	Pembahasan Hasil <i>Fine-tuning</i>	36
4.5	Evaluasi.....	37
4.5.1	Metrik Evaluasi	37
4.5.2	Hasil Perbandingan Model.....	37
4.6	Pembuatan Antarmuka	39
4.6.1	Hasil Uji Fungsional Antarmuka	44
4.7	Simulasi Sistem OCR.....	45
4.7.1	Hasil Simulasi Lima Citra Label Makanan.....	45
4.7.2	Hasil Simulasi Tiga Citra Label Obat	46
BAB V KESIMPULAN DAN SARAN		50
5.1	Kesimpulan	50
5.2	Saran.....	50
DAFTAR PUSTAKA.....		52
LAMPIRAN		55

DAFTAR TABEL

Tabel 2.1 Perbandingan penelitian sebelumnya.....	14
Tabel 2.2 Perbandingan model deteksi teks.....	16
Tabel 2.3 Perbandingan model pengenalan teks.....	16
Tabel 3.1 Konfigurasi hyperparameter utama <i>fine-tuning</i>	23
Tabel 4.1 Hasil evaluasi model pada data uji.	37
Tabel 4.2 Hasil evaluasi lima citra label makanan.	45
Tabel 4.3 Hasil evaluasi tiga citra label obat.	47
Tabel 4.4 Perbandingan citra dan prediksi.....	47

DAFTAR GAMBAR

Gambar 2.1 Arsitektur CRAFT.....	7
Gambar 2.2 Alur pengenalan teks menggunakan CRNN dengan kerangka TRBA	8
Gambar 2.3 Ilustrasi rektifikasi teks menggunakan TPS.....	9
Gambar 2.4 <i>Residual connection</i> pada blok ResNet.....	9
Gambar 2.5 Struktur BiLSTM	10
Gambar 2.6 Mekanisme attention pada tahap prediksi karakter.....	11
Gambar 3.1 Alur penelitian OCR untuk ekstraksi daftar bahan label makanan.....	18
Gambar 3.2 Data citra pada <i>dataset Ingredients Image from Food Label</i> dari Kaggle.....	20
Gambar 3.3 Alur sistem OCR.....	21
Gambar 3. 4 Alur <i>fine-tuning</i> model CRNN.....	22
Gambar 4.1 Contoh citra <i>dataset</i> utama label bahan makanan.....	26
Gambar 4.2 Hasil ekstraksi sub-citra berbasis kata pada contoh citra.....	27
Gambar 4.3 Pasangan label dan sub-citra setelah verifikasi.....	27
Gambar 4.4 Lima citra eksternal pada produk kemasan.....	28
Gambar 4.5 Tiga citra eksternal label obat.....	28
Gambar 4.6 Contoh citra yang akan dideteksi.....	29
Gambar 4.7 Hasil peta region teks pada citra deteksi.....	30
Gambar 4.8 Hasil peta keterhubungan antarkarakter pada citra deteksi.....	30
Gambar 4.9 Hasil <i>overlay</i> citra deteksi.....	31
Gambar 4.10 Hasil pengenalan pada level kata.....	33
Gambar 4.11 Gabungan per baris untuk baris ke-2 pada citra deteksi.....	33
Gambar 4.12 Gabungan per citra pada citra deteksi.....	34
Gambar 4.13 Contoh gabungan lintas 3 citra.....	34
Gambar 4.14 Kurva <i>training loss vs validation loss</i> saat <i>fine-tuning</i>	35
Gambar 4.15 Kurva <i>Word Accuracy</i> saat validasi <i>fine-tuning</i>	36
Gambar 4.16 Kurva <i>NED</i> saat validasi <i>fine-tuning</i>	36
Gambar 4.17 Tampilan area pengaturan inferensi.....	40
Gambar 4.18 Tampilan perintah yang dieksekusi.....	40
Gambar 4.19 Bagian penyuntingan citra sebelum inferensi.....	41
Gambar 4.20 Bagian penyuntingan potongan citra jika <i>checkbox</i> “Enable crop” diaktifkan..	42
Gambar 4.21 Tampilan bagian “Stdout”.....	42
Gambar 4.22 Tampilan bagian “Merged Output” dan “Input Images”.....	43

Gambar 4.23 Tampilan bagian “Recognized Table”	44
Gambar 4.24 Hasil simulasi sistem pada lima citra eksternal.	46

BAB I PENDAHULUAN

1.1 Latar Belakang

Informasi daftar bahan pada kemasan makanan umumnya masih tersaji sebagai teks cetak pada permukaan kemasan fisik. Dalam konteks sistem digital, teks tersebut belum langsung tersedia sebagai data yang dapat disimpan, dicari, atau diproses lebih lanjut karena masih berada dalam citra. Oleh karena itu, diperlukan proses ekstraksi teks dari citra kemasan menjadi teks digital agar informasi daftar bahan dapat digunakan sebagai keluaran terstruktur oleh sistem komputer. Secara umum, proses ini dapat dilakukan dengan *Optical Character Recognition* (OCR), yaitu teknologi untuk mendeteksi dan mengenali teks pada citra (Wang et al., 2023).

Ekstraksi teks pada label makanan memiliki tantangan visual yang berbeda dari dokumen teks biasa. Label kemasan dapat memuat teks berukuran kecil, padat, memiliki variasi jenis huruf, tersusun dalam tata letak yang tidak seragam, serta dipengaruhi kualitas citra seperti blur, pantulan cahaya, distorsi perspektif, dan permukaan kemasan yang melengkung. Penelitian mengenai *grocery label detection and recognition* menunjukkan bahwa variasi visual kemasan dan kompleksitas tata letak menjadi faktor penting yang memengaruhi keberhasilan deteksi serta pengenalan teks pada label produk (Guimarães et al., 2023). Evaluasi OCR pada label kemasan pangan juga menunjukkan bahwa citra dunia nyata masih menimbulkan tantangan karena teks sering muncul dalam susunan padat, kualitas gambar bervariasi, dan permukaan kemasan dapat menyebabkan gangguan visual (Nagayi et al., 2025).

Dalam konteks ini, OCR digunakan sebagai pendekatan teknis untuk mengubah teks pada citra kemasan menjadi teks digital. Pada pendekatan modern, proses OCR umumnya terdiri dari deteksi area teks dan pengenalan isi teks (Wang et al., 2023). Pemisahan dua tahap tersebut diperlukan karena sistem harus terlebih dahulu menemukan lokasi kata pada citra sebelum setiap potongan teks dapat dikenali dan disusun kembali menjadi keluaran teks yang utuh. Studi pengenalan tanggal kedaluwarsa pada kemasan pangan juga menunjukkan bahwa alur deteksi, pemotongan area teks, dan pengenalan karakter menjadi komponen penting dalam sistem OCR untuk citra kemasan (Peng et al., 2025). Penelitian tentang ekstraksi informasi label produk menunjukkan bahwa alur berbasis machine vision dan OCR dapat digunakan untuk mengotomatisasi pengambilan informasi tekstual dari label produk, tetapi kualitas hasil tetap

bergantung pada kemampuan sistem dalam mendeteksi dan mengenali teks secara tepat (Seitaj & Elangovan, 2024).

Meskipun model OCR pra-latih telah banyak tersedia, penerapannya secara langsung pada domain label makanan belum tentu menghasilkan kinerja yang stabil. Banyak model pengenalan teks dilatih dan dibandingkan menggunakan data scene text umum atau data sintesis berskala besar, sedangkan label bahan makanan memiliki karakter visual yang lebih spesifik, seperti teks kecil, kepadatan kata tinggi, dan susunan informasi yang dekat dengan permukaan kemasan (J. Baek et al., 2019; Nagayi et al., 2025). Dalam kajian *transfer learning*, perbedaan distribusi antara data sumber dan data target dapat mengurangi efektivitas pemindahan pengetahuan, terutama ketika fitur yang dipelajari semakin spesifik terhadap data asal (Pan & Yang, 2009; Yosinski et al., 2014). Oleh karena itu, penggunaan model pra-latih perlu diikuti dengan *fine-tuning* pada data label makanan agar representasi model dapat menyesuaikan pola visual dan karakter teks pada domain target.

Pada penelitian ini, tahap deteksi teks menggunakan *Character Region Awareness for Text Detection* (CRAFT), sedangkan tahap pengenalan teks menggunakan *Convolutional Recurrent Neural Network* (CRNN) dengan konfigurasi TPS-ResNet-BiLSTM-Attn (TRBA). CRAFT dipilih karena bekerja pada tingkat karakter melalui prediksi *region score* dan *affinity score*, sehingga sesuai untuk menangani teks alami yang memiliki variasi bentuk dan jarak antarkarakter (Y. Baek et al., 2019). Sementara itu, kerangka TRBA pada pengenalan teks menyediakan alur modular yang mencakup transformasi, ekstraksi fitur, pemodelan sekuens, dan prediksi karakter, sehingga banyak digunakan dalam kajian *scene text recognition* (J. Baek et al., 2019). Dengan dasar tersebut, *fine-tuning* pada CRNN-TRBA digunakan sebagai langkah adaptasi agar model pengenalan lebih sesuai dengan pola visual dan distribusi kata pada domain daftar bahan makanan (Nagayi et al., 2025; Yosinski et al., 2014).

Berdasarkan uraian tersebut, penelitian ini diarahkan untuk merancang dan mengevaluasi sistem OCR pada domain label bahan makanan. Sistem dikembangkan menggunakan CRAFT pada tahap deteksi teks dan CRNN-TRBA pada tahap pengenalan teks, kemudian model pengenalan disesuaikan melalui *fine-tuning* menggunakan data label makanan. Evaluasi dilakukan untuk menilai kemampuan sistem dalam mengekstraksi teks daftar bahan dari citra kemasan secara akurat.

1.2 Rumusan Masalah

Bahan pada label makanan merupakan informasi yang sangat penting dalam mendukung keputusan konsumsi, terutama bagi konsumen yang perlu memastikan kandungan bahan tertentu sebelum membeli atau mengonsumsi produk. Namun pada praktiknya, informasi tersebut masih sulit dibaca dan dipahami karena biasanya ditampilkan dalam ukuran huruf yang kecil, padat, menggunakan istilah teknis, dan sering kali ditulis dalam bahasa asing pada kemasan fisik. Selain itu, informasi komposisi yang masih berbentuk teks di dalam citra belum dapat langsung dimanfaatkan untuk kebutuhan pencarian, penyaringan bahan, maupun analisis lanjutan. Kondisi ini semakin diperumit oleh keterbatasan model OCR umum yang belum selalu stabil ketika diterapkan pada label makanan, mengingat adanya variasi kualitas citra, tata letak label, serta bentuk karakter teks yang beragam. Berdasarkan permasalahan tersebut, rumusan masalah dalam penelitian ini adalah bagaimana merancang dan mengevaluasi sistem OCR yang sesuai untuk domain label makanan sehingga teks bahan pada citra kemasan dapat diekstraksi menjadi teks digital secara akurat? Rumusan masalah ini menjadi dasar untuk menyusun langkah-langkah penelitian, merumuskan kebutuhan sistem, serta menilai sejauh mana pendekatan yang diusulkan mampu menjawab tantangan ekstraksi teks pada label makanan.

1.3 Batasan Penelitian

Penelitian ini difokuskan pada ruang lingkup berikut:

- a. Penelitian berfokus pada teks daftar bahan pada label makanan kemasan berbahasa Inggris.
- b. Teks yang diproses dibatasi pada teks cetak pada kemasan, tulisan tangan tidak termasuk cakupan penelitian.
- c. Masukan sistem dibatasi pada citra statis label makanan kemasan, bukan video atau pemrosesan *real-time*.
- d. Data utama penelitian menggunakan sumber terbuka yang relevan dengan konteks label bahan makanan.
- e. Penelitian menitikberatkan pada proses deteksi teks, pengenalan teks, dan penyusunan keluaran teks digital.
- f. Model deteksi CRAFT digunakan sebagai model pra-latih tanpa *fine-tuning* ulang.
- g. *Fine-tuning* dilakukan pada model pengenalan CRNN-TRBA menggunakan data domain label bahan makanan.

- h. Evaluasi dilakukan pada data uji internal menggunakan metrik *Word Accuracy*, *Normalized Edit Distance (NED)*, *Character Error Rate (CER)*, dan *Word Error Rate (WER)*.
- i. Penelitian tidak membahas klasifikasi bahan halal dan non-halal, serta tidak membahas klasifikasi alergen sebagai keluaran utama sistem.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk merancang dan mengembangkan sistem OCR berbasis pembelajaran mendalam yang mampu mendeteksi dan mengenali teks daftar bahan pada label makanan kemasan secara akurat. Untuk mencapai tujuan tersebut, penelitian ini difokuskan pada perancangan alur sistem OCR untuk ekstraksi teks, penyesuaian model melalui proses fine-tuning menggunakan data domain label makanan, serta evaluasi kinerja sistem menggunakan metrik evaluasi OCR. Selain itu, penelitian ini juga bertujuan untuk menganalisis pola kesalahan pengenalan guna memahami pengaruh penyesuaian model terhadap kualitas hasil ekstraksi teks.

1.5 Manfaat Penelitian

Secara praktis, penelitian ini memberikan fondasi bagi pengembangan sistem yang mampu membantu pembacaan komposisi bahan pada label makanan kemasan secara lebih cepat, akurat, dan terstruktur. Sistem ini diharapkan dapat mendukung pengembangan layanan lanjutan berbasis teks komposisi, seperti pencarian informasi bahan, penyaringan kandungan tertentu, serta identifikasi bahan berisiko bagi kelompok konsumen tertentu. Secara teoretis, penelitian ini memberikan kontribusi dalam pengembangan kajian OCR, khususnya pada penerapan metode pembelajaran mendalam untuk ekstraksi teks pada domain label makanan. Selain itu, penelitian ini juga menambah bukti empiris terkait efektivitas penyesuaian model dengan data domain spesifik dalam meningkatkan kinerja sistem OCR.

1.6 Sistematika Penulisan

Struktur laporan penelitian ini disusun sebagai berikut:

- a. Bab I Pendahuluan, menyajikan latar belakang, rumusan masalah dan pertanyaan penelitian, batasan penelitian, tujuan dan manfaat penelitian, metodologi penelitian secara umum, serta sistematika penulisan.

- b. Bab II Kajian Pustaka, menyajikan dasar teori yang meliputi OCR, CRAFT, CRNN-TRBA, *fine-tuning*, metrik evaluasi, LMDB, Streamlit, kemudian dilanjutkan dengan kajian penelitian sebelumnya yang relevan dan pemilihan metode.
- c. Bab III Metodologi Penelitian, menyajikan langkah-langkah penelitian dan uraian penerapan metode, mencakup pengumpulan data, preprocessing, pembuatan sistem OCR, *fine-tuning* model pengenalan teks, evaluasi pada data uji, perancangan antarmuka, dan simulasi sistem OCR.
- d. Bab IV Hasil dan Pembahasan, menyajikan hasil implementasi dan pembahasan pada setiap tahapan penelitian, meliputi pengumpulan data, preprocessing, implementasi sistem OCR, hasil *fine-tuning*, evaluasi model, simulasi sistem OCR, serta hasil uji fungsional antarmuka.
- e. Bab V Kesimpulan dan Saran, menyajikan kesimpulan penelitian dan saran pengembangan lanjutan.

BAB II

DAFTAR PUSTAKA

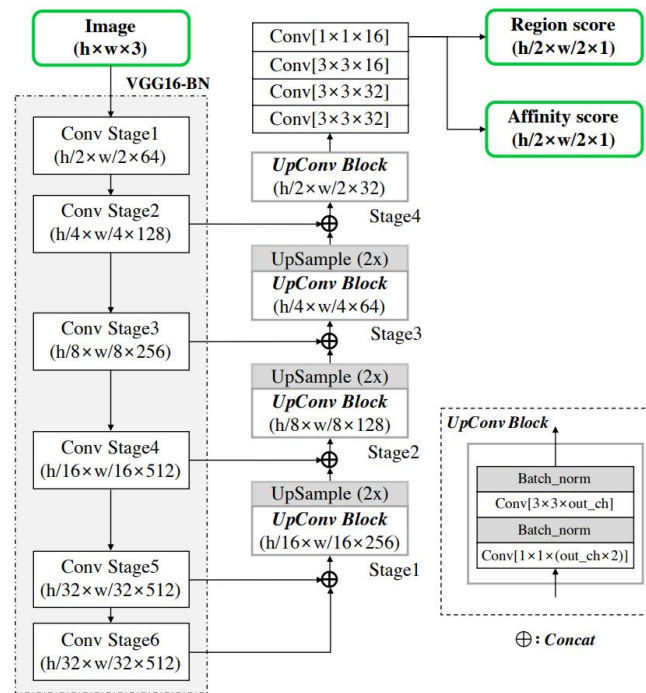
2.1 Dasar Teori

2.1.1 *Optical Character Recognition (OCR)*

OCR adalah proses mengubah teks pada citra menjadi teks digital yang dapat dicari dan diolah lebih lanjut. Pada pendekatan modern, OCR umumnya dibagi menjadi dua tahap inti, yaitu *text detection* (deteksi area teks) dan *text recognition* (pengenalan teks) (Wang et al., 2023). Untuk konteks label pangan, tantangan utama bukan hanya keberadaan teks, tetapi juga variasi ukuran huruf, kepadatan informasi, latar kemasan, dan gangguan visual yang memengaruhi ketepatan pembacaan (Guimarães et al., 2023; Seitaj & Elangovan, 2024). Oleh karena itu, rancangan OCR pada penelitian ini menekankan kombinasi metode deteksi dan pengenalan teks yang mampu menangani variasi tersebut, yaitu dengan memanfaatkan pendekatan berbasis CRAFT untuk deteksi area teks dan CRNN-TRBA untuk pengenalan teks.

2.1.2 *Character Region Awareness for Text detection (CRAFT)*

CRAFT merupakan metode deteksi teks yang bekerja pada level karakter, sehingga lebih sensitif terhadap teks kecil, rapat, dan tidak selalu seragam pada citra kemasan (Y. Baek et al., 2019). Secara konseptual, CRAFT membentuk dua peta prediksi, yaitu *region score* (peta wilayah karakter) dan *affinity score* (peta keterhubungan antarkarakter), kemudian kedua peta tersebut digabungkan untuk memperoleh area kata. Arsitektur CRAFT menggunakan *backbone* konvolusional berbasis VGG-16 yang dipadukan dengan skema upsampling untuk mempertahankan detail spasial pada proses lokalisasi teks (Y. Baek et al., 2019). Karakteristik ini relevan untuk domain label pangan yang memiliki kepadatan informasi tinggi dan variasi latar visual (Guimarães et al., 2023; Wang et al., 2023). Arsitektur CRAFT yang digunakan pada penelitian ini ditunjukkan pada Gambar 2.1.



Gambar 2.1 Arsitektur CRAFT

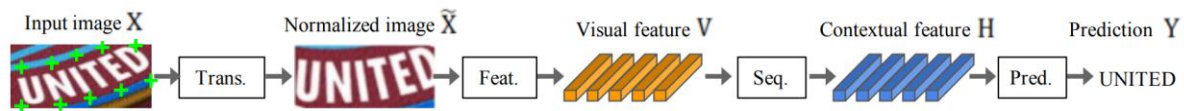
Sumber: (Y. Baek et al., 2019)

Pada model CRAFT, dataset yang digunakan mencakup SynthText serta dataset benchmark ICDAR, yaitu IC13, IC15, dan IC17. SynthText adalah dataset sintetis yang menempatkan teks buatan ke dalam citra pemandangan alami, sehingga model terpapar variasi font, ukuran, orientasi, dan kondisi latar secara luas (Gupta et al., 2016). Sementara itu, IC13, IC15, dan IC17 berisi citra teks alami dengan anotasi lokasi teks pada kondisi lapangan yang beragam (Y. Baek et al., 2019). Kombinasi data sintetis dan data nyata tersebut mendukung pembentukan kemampuan deteksi CRAFT pada berbagai pola teks dan variasi visual kemasan.

2.1.3 Convolution Recurrent Neural Network (CRNN) dengan Kerangka TPS-ResNet-BiLSTM-Attention (TRBA)

CRNN merupakan arsitektur pengenalan teks yang menggabungkan ekstraksi fitur visual, pemodelan urutan, dan prediksi karakter dalam satu alur pengenalan (Shi et al., 2015). Pada implementasi penelitian ini, CRNN ditempatkan sebagai modul pengenalan teks setelah tahap deteksi area teks oleh CRAFT. Konfigurasi yang digunakan mengikuti kerangka TRBA, yaitu transformasi (Trans.) dengan TPS, ekstraksi fitur (Feat.) dengan ResNet, pemodelan sekuens (Seq.) dengan BiLSTM, dan prediksi (Pred.) dengan Attention (J. Baek et al., 2019). Pendekatan ini dipilih karena bersifat modular, sehingga adaptasi pada domain label pangan

dapat dilakukan secara terarah tanpa mengubah prinsip dasar alur OCR (J. Baek et al., 2019; Guimarães et al., 2023). Alur pengenalan teks menggunakan CRNN dengan kerangka TRBA ditunjukkan pada Gambar 2.2.



Gambar 2.2 Alur pengenalan teks menggunakan CRNN dengan kerangka TRBA

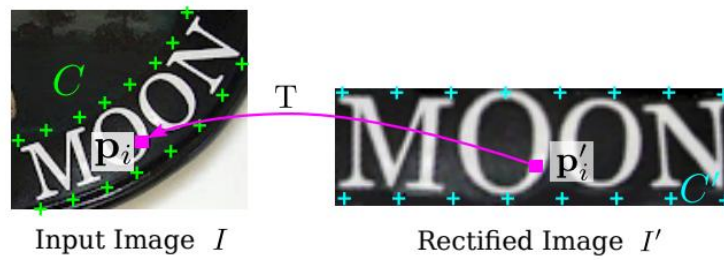
Sumber: (J. Baek et al., 2019)

Model CRNN-TRBA menggunakan kombinasi dataset MJSynth (MJ) dan SynthText (ST) untuk pelatihan (J. Baek et al., 2019). MJSynth adalah dataset kata sintetis skala besar untuk scene text recognition yang dibangun dari proses perenderan kata dengan variasi font, warna, efek, dan distorsi sehingga menghasilkan keragaman bentuk karakter (Jaderberg et al., 2014). SynthText melengkapi variasi tersebut dengan menempatkan teks sintetis ke citra alamiah, sehingga model mempelajari karakter pada latar visual yang lebih kompleks (Gupta et al., 2016). Kombinasi MJ dan ST tersebut menyediakan cakupan data yang luas untuk pembelajaran awal pengenalan karakter sebelum model disesuaikan lebih lanjut pada data penelitian.

Untuk memperjelas alur kerja CRNN-TRBA pada penelitian ini, empat komponen utamanya dijabarkan secara bertahap, mulai dari TPS, ResNet, BiLSTM, dan Attention.

A. *Thin Plate Spline* (TPS)

TPS adalah modul transformasi geometri yang berfungsi melakukan rektifikasi citra teks sebelum tahap ekstraksi fitur. Modul ini membantu menormalkan teks yang melengkung atau terdistorsi perspektif, sehingga bentuk karakter lebih konsisten saat diproses tahap berikutnya (Shi et al., 2016). Dalam konteks label kemasan, peran TPS penting karena posisi kamera dan permukaan kemasan dapat menyebabkan kemiringan teks yang signifikan. Ilustrasi proses rektifikasi teks menggunakan TPS ditunjukkan pada Gambar 2.3.

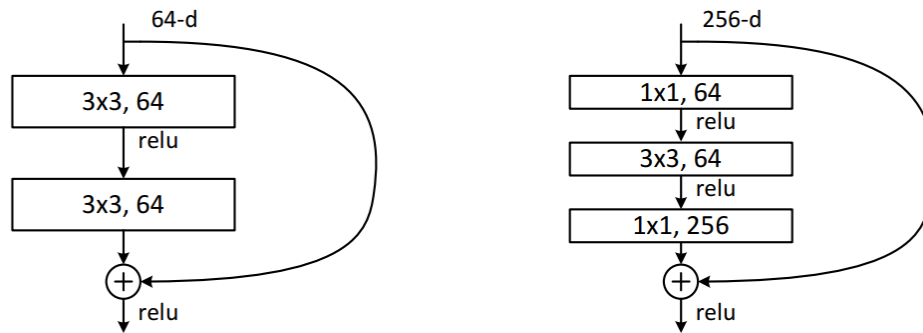


Gambar 2.3 Ilustrasi rektifikasi teks menggunakan TPS

Sumber: (Shi et al., 2016)

B. *Residual Network (ResNet)*

ResNet berfungsi sebagai pengekstrak fitur visual pada tahap pengenalan. Mekanisme *residual connection* memungkinkan jaringan lebih dalam tetap stabil saat pelatihan, sehingga representasi fitur karakter menjadi lebih kuat dibanding arsitektur konvolusional yang lebih dangkal (He et al., 2015). Pada kerangka TRBA, ResNet digunakan untuk menghasilkan fitur yang kaya terhadap variasi bentuk huruf, ketajaman citra, dan gangguan latar, yang merupakan karakter umum pada citra label produk. Skema *residual connection* pada blok ResNet ditunjukkan pada Gambar 2.4.



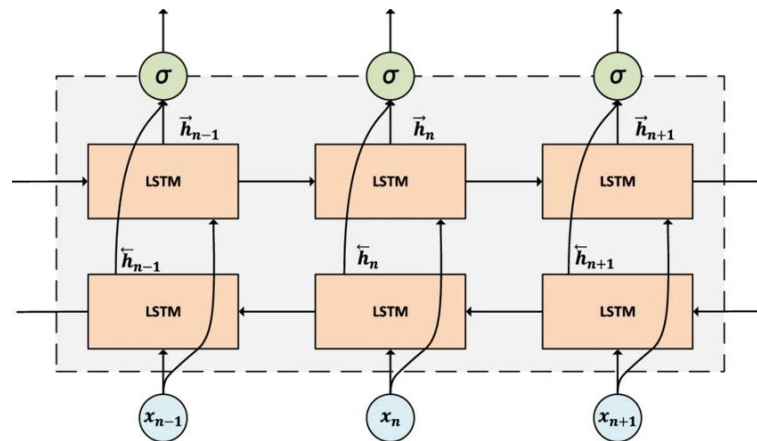
Gambar 2.4 *Residual connection* pada blok ResNet

Sumber: (He et al., 2015)

C. *Bidirectional Long Short-Term Memory (BiLSTM)*

BiLSTM digunakan untuk memodelkan konteks urutan karakter dari 2 arah, yaitu kiri ke kanan dan kanan ke kiri. Pendekatan dua arah ini membantu model memahami ketergantungan antarkarakter dalam kata secara lebih lengkap, terutama pada kata panjang atau pola ejaan yang saling mirip (J. Baek et al., 2019; Shi et al., 2015). Dengan demikian, BiLSTM berperan

sebagai penguat stabilitas pembacaan urutan karakter sebelum prediksi akhir dilakukan. Struktur BiLSTM yang memodelkan konteks dua arah ditunjukkan pada Gambar 2.5.

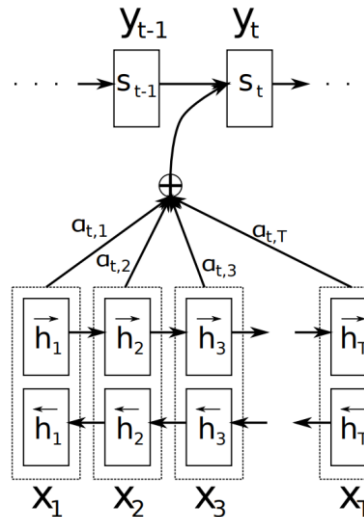


Gambar 2.5 Sruktur BiLSTM

Sumber: (Magadán et al., 2024)

D. Attention

Attention adalah mekanisme prediksi yang memberi bobot fokus pada bagian fitur yang paling relevan saat menghasilkan karakter demi karakter. Pada pengenalan teks, mekanisme ini membantu model menentukan posisi informasi visual yang tepat untuk setiap langkah prediksi, sehingga kesalahan urutan karakter dapat dikurangi pada teks yang tidak teratur (J. Baek et al., 2019). Kajian OCR terbaru juga menunjukkan bahwa mekanisme perhatian tetap menjadi komponen penting, baik pada arsitektur berbasis rekuren maupun *transformer* (Li et al., 2021; Neudecker et al., 2021). Mekanisme attention pada tahap prediksi karakter ditunjukkan pada Gambar 2.6.



Gambar 2.6 Mekanisme attention pada tahap prediksi karakter

Sumber: (Bahdanau et al., 2014)

2.1.4 *Fine-tuning*

Fine-tuning merupakan proses adaptasi bobot model pra-latih menggunakan data domain target agar representasi fitur lebih sesuai dengan karakteristik data penelitian. Pada kasus label pangan, proses ini diperlukan karena istilah bahan, kepadatan teks, dan kualitas citra kemasan berbeda dari data teks alami umum yang biasanya digunakan saat pra-pelatihan model (Guimarães et al., 2023; Nagayi et al., 2025). Tujuan utamanya adalah meningkatkan ketepatan pengenalan pada domain spesifik tanpa membangun model dari awal.

2.1.5 **Metrik Evaluasi**

Evaluasi pada penelitian OCR tidak cukup dilihat dari satu metrik saja, karena tiap metrik menangkap aspek kesalahan yang berbeda. Pada penelitian ini digunakan empat metrik, yaitu *Word Accuracy*, Normalized Edit Distance (NED), *Character Error Rate* (CER), dan *Word Error Rate* (WER) (Neudecker et al., 2021; Wang et al., 2023).

A. *Word Accuracy*

Word Accuracy mengukur persentase prediksi kata yang tepat sama persis dengan label acuan. Metrik ini digunakan luas pada evaluasi pengenalan teks karena langsung menunjukkan berapa banyak kata yang dikenali tanpa kesalahan (J. Baek et al., 2019). Nilai *Word Accuracy* dihitung dari rentang 0% hingga 100%. Semakin tinggi nilai *Word Accuracy*, semakin banyak

kata yang dikenali secara tepat. Namun, metrik ini bersifat ketat karena satu kesalahan karakter saja akan membuat satu kata dihitung salah. Metrik ini ditunjukkan pada Persamaan 2.1.

$$Word Accuracy = \frac{N_{correct}}{N_{total}} \times 100\% \quad (2.1)$$

Keterangan:

$N_{correct}$ = jumlah kata yang prediksinya identik dengan *ground truth*.

N_{total} = jumlah seluruh kata yang diuji.

B. Normalized Edit Distance (NED)

NED mengukur tingkat kemiripan prediksi terhadap *ground truth* berdasarkan jarak edit Levenshtein yang dinormalisasi (J. Baek et al., 2019; Neudecker et al., 2021). Berbeda dengan *Word Accuracy* yang bersifat biner benar-salah, NED tetap memberi skor parsial untuk prediksi yang hampir benar. Nilai NED berada pada rentang 0 sampai 1. Nilai yang semakin mendekati 1 menunjukkan prediksi semakin mirip dengan label acuan. Metrik ini ditunjukkan pada Persamaan 2.2.

$$NED = 1 - \frac{D(P,G)}{\max(|P|,|G|)} \quad (2.2)$$

Keterangan:

$D(P, G)$ = edit distance (Levenshtein distance), antar teks prediksi P dan teks *ground truth* G , yaitu jumlah minimum operasi edit (substitusi, penyisipan, dan penghapusan karakter) yang dibutuhkan untuk mengubah P menjadi G .

P = teks hasil prediksi model.

G = teks *ground truth*.

$|P|$ = panjang karakter dari teks prediksi.

$|G|$ = panjang karakter dari teks *ground truth*.

C. Character Error Rate (CER)

CER mengukur proporsi kesalahan pada level karakter, sehingga cocok untuk menganalisis kesalahan detail pada istilah bahan pangan (Nagayi et al., 2025; Neudecker et al., 2021). Metrik ini sensitif terhadap salah eja kecil yang mungkin tidak terlihat pada metrik akurasi saja. Nilai CER berada pada rentang 0 sampai 1. Semakin kecil nilainya menunjukkan kualitas pengenalan karakter yang lebih baik. Metrik ini ditunjukkan pada Persamaan 2.3.

$$CER = \frac{S + D + I}{N} \quad (2.3)$$

Keterangan:

S = jumlah karakter substitusi.

D = jumlah karakter terhapus.

I = jumlah karakter tersisip.

N = total karakter pada *ground truth*.

D. Word Error Rate (WER)

WER mengukur kesalahan pada level kata, sehingga berguna untuk menilai keterbacaan hasil OCR sebagai satuan kata utuh (Nagayi et al., 2025; Neudecker et al., 2021). Metrik ini relevan pada ekstraksi daftar bahan karena kesalahan satu kata dapat memengaruhi interpretasi isi label. Nilai WER berada pada rentang 0 sampai 1. Semakin kecil nilainya menunjukkan kualitas rekonstruksi kata yang lebih baik. Metrik ini ditunjukkan pada Persamaan 2.4.

$$WER = \frac{S+D+I}{N_w} \quad (2.4)$$

Keterangan:

S = jumlah kata substitusi.

D = jumlah kata terhapus.

I = jumlah kata tersisip.

N_w = total kata pada *ground truth*.

2.1.6 Lightning Memory-Mapped Database (LMDB)

LMDB adalah basis data key-value tersemat yang menggunakan *memory-mapped file* sehingga data dapat diakses langsung dari ruang alamat proses tanpa lapisan salinan *buffer* yang berulang. Secara arsitektural, LMDB menggabungkan struktur B+tree, mekanisme *copy-on-write*, dan *multiversion concurrency control* (MVCC), sehingga banyak transaksi baca dapat berjalan konsisten bersamaan dengan transaksi tulis tunggal. Karakteristik ini relevan untuk pipeline pelatihan OCR karena pasangan sub-citra dan label dapat disimpan serta dibaca ulang secara efisien saat proses batching data. Dalam konteks penelitian ini, penggunaan LMDB mendukung alur pelatihan yang stabil ketika jumlah sampel cukup besar dan akses baca dilakukan berulang pada setiap iterasi (Chu, 2011).

2.1.7 Streamlit

Streamlit merupakan *framework* Python yang banyak digunakan untuk membangun antarmuka web pada aplikasi *machine learning*, terutama untuk menjembatani model dengan pengguna nonteknis melalui alur input data, eksekusi model, dan penayangan hasil prediksi (Buga et al., 2025). Dalam penelitian ini, Streamlit digunakan untuk merancang antarmuka inferensi OCR yang mencakup pengaturan parameter, menjalankan proses OCR, serta menampilkan keluaran secara terstruktur. Penggunaan Streamlit pada berbagai penelitian menunjukkan bahwa pendekatan ini sesuai untuk kebutuhan purwarupa karena implementasi antarmuka dapat diintegrasikan langsung dengan *pipeline* Python tanpa pemisahan pengembangan *front-end* yang kompleks (Buga et al., 2025). Karakteristik tersebut mendukung kebutuhan penelitian ini yang menuntut iterasi pengujian konfigurasi inferensi secara cepat pada satu antarmuka.

2.2 Penelitian Sebelumnya

Ringkasan penelitian sebelumnya yang menjadi rujukan disajikan pada Tabel 2.1.

Tabel 2.1 Perbandingan penelitian sebelumnya

Penulis	Judul Penelitian	Model	Fokus	Hasil
(Rosyadi et al., 2024)	<i>Ingredients Identification Through Label Scanning Using PaddleOCR and ChatGPT</i>	PaddleOCR (DB + SVTR_LCNet)	OCR pada label komposisi bahan makanan kemasan melalui pemindaian <i>smartphone</i> .	CER rata-rata 0,14. Performa menurun pada kemasan reflektif dan bertekstur.
(Manuel et al., 2024)	<i>Implementation of Tesseract OCR and Bounding Box for Text Extraction on Food Nutrition Labels</i>	MobileNetV1 dan Tesseract	Ekstraksi teks label nutrisi makanan dengan deteksi area sebelum OCR.	WER 10%, dan CER 1,6%. Pemisahan tahap deteksi dan pengenalan meningkatkan akurasi secara signifikan.
(Seitaj & Elangovan, 2024)	<i>Information Extraction from Product Labels: A Machine Vision Approach</i>	CRNN, Tesseract, dan NLP	Ekstraksi teks label produk makanan menggunakan CRNN untuk menangani variasi <i>font</i> dan desain label.	CRNN efektif pada label makanan dan mampu mengenali baris teks panjang tanpa segmentasi per karakter.

Berdasarkan Tabel 2.3, terdapat tiga penelitian yang secara langsung relevan sebagai landasan pemilihan model CRAFT dan CRNN dalam penelitian ini. (Rosyadi et al., 2024) menjadi rujukan utama dari sisi kesesuaian domain karena secara spesifik berfokus pada pemindaian label komposisi bahan makanan menggunakan *smartphone*. Meskipun menggunakan PaddleOCR, penelitian tersebut menunjukkan adanya penurunan performa pada kemasan dengan permukaan reflektif dan bertekstur, dengan nilai CER rata-rata sebesar 0,14. Temuan ini mengindikasikan bahwa tahap deteksi teks memiliki peran penting dalam menjaga akurasi sistem, terutama pada kondisi citra yang tidak ideal.

Pentingnya pemisahan tahap deteksi dan pengenalan teks juga telah ditunjukkan melalui penggunaan deteksi area berbasis *bounding box* sebelum proses pengenalan yang mampu menurunkan CER hingga 1,6% pada label nutrisi (Manuel et al., 2024). Hasil ini menunjukkan bahwa pendekatan dua tahap memberikan kinerja yang lebih baik dibandingkan pendekatan OCR satu tahap.

Dari sisi arsitektur pengenalan teks, CRNN mampu menangani variasi *font* dan tata letak teks pada label produk makanan melalui kombinasi ekstraksi fitur visual dan pemodelan urutan karakter (Seitaj & Elangovan, 2024). Kemampuan ini menjadikan CRNN sesuai untuk mengenali teks panjang dan padat tanpa perlu segmentasi karakter secara eksplisit. Selain itu, fleksibilitas arsitektur CRNN dalam memproses urutan karakter juga memungkinkan model ini beradaptasi dengan berbagai bentuk teks yang tidak terstruktur pada label kemasan makanan.

Berdasarkan berbagai penelitian tersebut, dapat disimpulkan bahwa keberhasilan OCR pada label makanan sangat dipengaruhi oleh kemampuan deteksi teks serta arsitektur pengenalan yang mampu menangani teks padat dan bervariasi. Namun, beberapa penelitian masih menggunakan model umum tanpa penyesuaian khusus terhadap karakteristik domain label makanan, sehingga masih terdapat peluang untuk meningkatkan kinerja sistem melalui pendekatan yang lebih adaptif terhadap domain spesifik yaitu daftar bahan pada label makanan.

2.3 Pemilihan Metode

Berdasarkan hasil analisis pada subbab penelitian sebelumnya, diperlukan pendekatan yang mampu memisahkan proses deteksi dan pengenalan teks serta menggunakan model yang memiliki kinerja tinggi pada masing-masing tahap tersebut. Pada penelitian ini, metode deteksi teks yang digunakan adalah CRAFT dan metode pengenalan teks yang digunakan adalah CRNN dengan arsitektur TRBA. Kombinasi CRAFT dan CRNN-TRBA dipilih karena keduanya saling melengkapi dalam membentuk *pipeline* OCR. CRAFT berperan dalam

memastikan setiap area teks terdeteksi secara akurat, sementara CRNN-TRBA bertugas mengenali karakter pada area tersebut secara kontekstual. Kombinasi ini diharapkan mampu meningkatkan kinerja sistem OCR dalam mengekstraksi teks komposisi pada label makanan, terutama pada kondisi citra yang kompleks dan bervariasi. Perbandingan kinerja model deteksi teks dan pengenalan teks yang menjadi dasar pemilihan CRAFT dan CRNN disajikan pada Tabel 2.2 dan Tabel 2.3.

Tabel 2.2 Perbandingan model deteksi teks

Sumber: (Y. Baek et al., 2019)

Model	MSRA-TD500			FPS
	Recall	Precision	H-mean	
EAST	67.4	87.3	76.1	13.2
TextSnake	73.9	83.2	78.3	1.1
PixelLink	73.2	83.0	77.8	3.0
RRD	73	87	79	10
Lyu et al.	76.2	87.6	81.5	5.7
CRAFT	93.1	97.4	95.2	8.6

Berdasarkan Tabel 2.1, CRAFT menunjukkan kinerja tertinggi pada seluruh metrik deteksi yang dibandingkan pada *benchmark* MSRA-TD500, yaitu *recall* 93,1, *precision* 97,4, dan *H-mean* 95,2. Dibandingkan model pembanding. Dari sisi kecepatan, CRAFT memang bukan yang paling cepat dibanding EAST, tetapi tetap memberikan kompromi pada akurasi yang jauh lebih tinggi. Oleh karena itu, CRAFT dipilih karena penelitian ini memprioritaskan ketelitian lokalisasi teks pada label makanan yang padat daripada kecepatan maksimum.

Tabel 2.3 Perbandingan model pengenalan teks

Sumber: (J. Baek et al., 2019)

Model	Data pelatihan	IIIT	SVT	SP	CT	Waktu ms/gambar
CRNN	MJ+ST	82.9	81.6	70.0	65.5	4.4
RARE	MJ+ST	86.2	85.8	76.2	70.4	23.6
R2AM	MJ+ST	83.4	82.4	72.1	64.9	24.1
STAR-Net	MJ+ST	87.0	86.9	77.5	71.7	10.9
GRCNN	MJ+ST	84.2	83.7	73.6	68.1	10.7
Rosetta	MJ+ST	84.3	84.7	73.8	69.2	4.7
CRNN-TRBA	MJ+ST	87.9	87.5	79.2	74.0	27.6

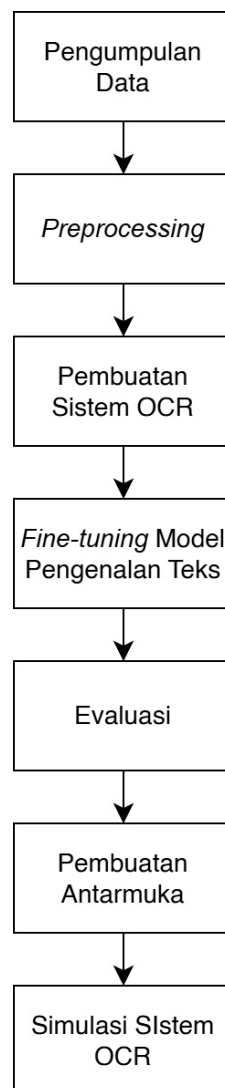
Berdasarkan Tabel 2.2, CRNN-TRBA menunjukkan kinerja paling tinggi pada empat benchmark utama, yaitu IIIT dengan nilai 87,9, SVT dengan nilai 87,5, SP dengan nilai 79,2,

dan CT dengan nilai 74,0. Dibandingkan CRNN dasar, peningkatan absolut CRNN-TRBA mencapai 5,0 poin pada IIIT, 5,9 poin pada SVT, 9,2 poin pada SP, dan 8,5 poin pada CT. Walaupun waktu inferensi CRNN-TRBA paling besar dibanding model lainnya, pola peningkatan akurasi yang konsisten pada seluruh benchmark menjadi dasar pemilihan CRNN-TRBA sebagai model pengenalan teks pada penelitian ini.

BAB III METODOLOGI PENELITIAN

3.1 Langkah-langkah Penelitian

Gambar 3.1 menunjukkan alur yang akan dilakukan pada penelitian ini. Langkah-langkah tersebut terdiri dari pengumpulan data, *preprocessing*, pembuatan sistem OCR, *fine-tuning* model pengenalan teks, evaluasi, dan simulasi sistem OCR.



Gambar 3.1 Alur penelitian OCR untuk ekstraksi daftar bahan label makanan

Berikut ini merupakan uraian tahapan secara berurutan mulai dari pengumpulan data hingga simulasi sistem OCR.

3.1.1 Pengumpulan Data

Pada penelitian ini, sumber data utama yang digunakan merujuk pada *dataset Ingredients Image from Food Label* yang tersedia pada platform Kaggle (*Ingredients Image from Food Label*, n.d.) dan bisa diakses pada <https://www.kaggle.com/datasets/shensivam/ingredients-image-from-food-label>. *Dataset* tersebut dipilih karena memuat 182 citra label bahan makanan berbahasa Inggris yang sesuai dengan konteks penelitian dan bisa digunakan untuk *fine-tuning*. 182 Citra label tersebut akan dipotong per kata menjadi sub-citra dan diberi label sesuai dengan teks pada sub-citra tersebut. Tahap pemotongan dan pelabelan dilakukan dengan cara:

- a. Pemotongan citra dilakukan dengan bantuan CRAFT dengan memotong seluruh kata yang ada pada citra menjadi sub-citra.
- b. Pelabelan akan dilakukan pada sub-citra dengan bantuan CRNN.
- c. Kesesuaian hasil sub-citra dan label dicek secara manual oleh penulis dengan menghapus sub-citra yang tidak terbaca katanya, dan mengubah label yang tidak sesuai dengan kata pada sub-citranya.

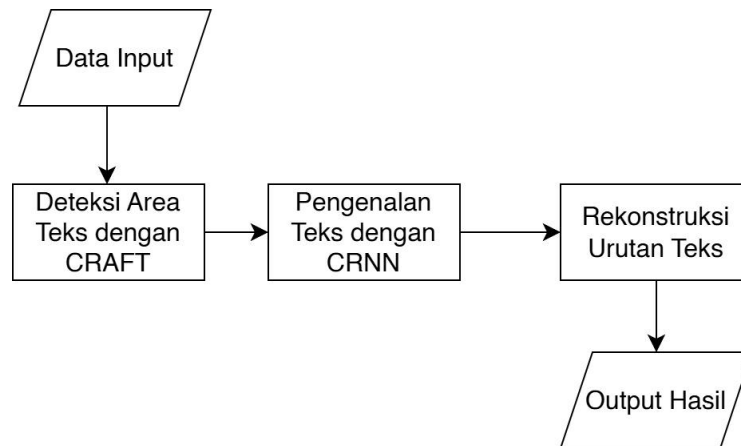
Secara konseptual, penggunaan *dataset* domain spesifik membantu model mempelajari variasi visual yang umum muncul pada label kemasan, seperti ukuran teks kecil, kepadatan kata, latar kemasan berwarna, dan kualitas citra yang tidak selalu seragam. Karakteristik tersebut penting karena sering menjadi sumber penurunan performa ketika model OCR yang dilatih pada data umum diterapkan pada domain pangan (Guimarães et al., 2023; Nagayi et al., 2025). Contoh data citra pada dataset *Ingredients Image from Food Label* dari Kaggle ditunjukkan pada Gambar 3.2.

Selain data utama, penelitian juga akan menggunakan data eksternal berupa citra label makanan di luar *dataset* utama untuk simulasi penggunaan sistem pada kondisi nyata.

Setelah proses pengumpulan data menghasilkan pasangan sub-citra dan label teks, *preprocessing* dilakukan dalam beberapa langkah untuk menyiapkan data ke format pelatihan model. Pertama, pasangan sub-citra dan label yang telah dikurasi disusun ke file anotasi berbasis teks. Kedua, data sub-citra didistribusikan ke tiga subset, yaitu pelatihan, validasi, dan pengujian dengan target rasio 80:10:10. Proporsi pembagian dapat tidak tepat secara matematis karena mempertimbangkan keterkaitan sub-citra yang berasal dari citra induk yang sama agar tidak terjadi kebocoran data antarsubset. Ketiga, file anotasi pada tiap subset dikonversi ke format *Lightning Memory-Mapped Database* (LMDB) agar kompatibel dengan *loader* pelatihan pada *pipeline* CRNN. Hasil konversi inilah yang akan digunakan sebagai masukan pada tahap *fine-tuning*.

Sistem OCR pada penelitian ini didesain sebagai alur *end-to-end* untuk mengubah citra label bahan makanan menjadi teks digital. Alur disusun agar setiap tahapan memiliki masukan dan keluaran yang jelas yang ditunjukkan pada Gambar 3.3. Secara umum, sistem terdiri dari

input citra, deteksi dan ekstraksi area teks, pengenalan teks, rekonstruksi urutan teks, dan keluaran akhir.



Gambar 3.3 Alur sistem OCR

Tahap pertama adalah input citra. Citra label makanan dimasukkan ke sistem dengan mempertahankan struktur tata letak aslinya sebagai konteks pembacaan. Pada tahap ini, sistem menyiapkan citra agar dapat diproses oleh modul deteksi teks tanpa mengubah urutan informasi pada label.

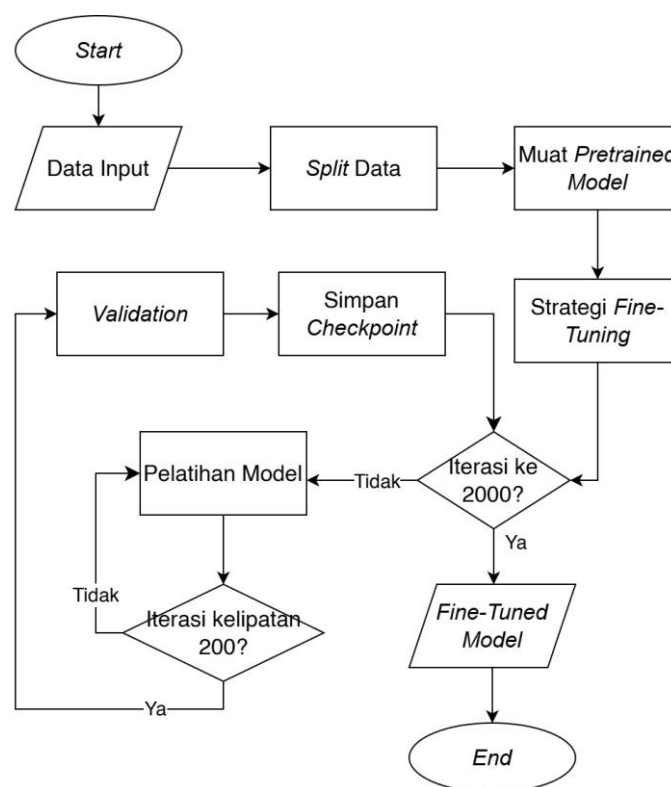
Tahap kedua adalah deteksi dan ekstraksi area teks menggunakan CRAFT. Model CRAFT akan menghasilkan *region score map* untuk menunjukkan probabilitas area karakter dan *affinity score map* untuk menunjukkan keterhubungan antar karakter. Kedua peta ini digunakan untuk membentuk kandidat *bounding box* pada level kata. Setiap area kata kemudian dipotong menjadi sub-citra melalui proses *word-level cropping* dengan transformasi perspektif agar teks yang miring tetap terbaca.

Tahap ketiga adalah pengenalan teks menggunakan CRNN dengan arsitektur TPS-ResNet-BiLSTM-Attention (TRBA). Modul TPS digunakan untuk menormalkan bentuk geometri teks, ResNet mengekstraksi fitur visual, BiLSTM memodelkan konteks urutan karakter dua arah, dan Attention melakukan *decoding* urutan karakter menjadi teks. Keluaran tahap ini berupa prediksi kata untuk setiap sub-citra beserta skor keyakinan prediksi.

Tahap keempat adalah rekonstruksi urutan teks. Prediksi pada level kata disusun kembali sesuai posisi spasial pada citra label, dengan aturan keterurutan dari kiri ke kanan dalam satu baris dan dari atas ke bawah antarbaris. Tahap ini bertujuan untuk membentuk kembali struktur daftar bahan agar mendekati susunan aslinya pada kemasan yang akhirnya menjadi keluaran pada tahap akhir sistem.

3.1.4 *Fine-tuning* Model Pengenalan Teks

Fine-tuning model pengenalan teks pada penelitian ini dirancang dengan memanfaatkan bobot pra-latih CRNN berarsitektur TPS-ResNet-BiLSTM-Attention (TRBA). Bobot awal CRNN berasal dari pelatihan pada MJSynth dan SynthText yang digunakan sebagai titik awal sebelum diadaptasi ke domain teks daftar bahan makanan. Berbeda dengan modul pengenalan, model CRAFT digunakan dalam bentuk bobot pra-latih untuk deteksi area teks dan tidak dilakukan *fine-tuning*. Alur *fine-tuning* model CRNN ditunjukkan pada Gambar 3. 4.



Gambar 3. 4 Alur *fine-tuning* model CRNN

Setelah bobot pra-latih dimuat, dilakukan strategi *fine-tuning* dengan membekukan parameter pada modul TPS dan ResNet. Hal ini dilakukan karena representasi pada lapisan awal cenderung lebih umum dan dapat dimanfaatkan kembali pada domain target, sedangkan lapisan yang lebih dekat ke keluaran cenderung lebih spesifik terhadap domain target yang dalam konteks ini berhubungan dengan teks pada label makanan. Karena itu, pembekuan TPS dan ResNet dilakukan untuk mempertahankan kemampuan normalisasi geometri serta fitur visual dasar yang sudah stabil. Pada saat yang sama, modul BiLSTM dan Attention tetap dilatih karena kedua modul tersebut berperan langsung pada pemodelan konteks urutan karakter dan

proses decoding teks yang perlu menyesuaikan distribusi data target. Strategi ini membantu model beradaptasi pada pola sekuens teks data target tanpa mengubah berlebihan komponen visual dasar yang telah stabil (Pan & Yang, 2009; Yosinski et al., 2014). Untuk konfigurasi hyperparameter utama pada proses *fine-tuning* disajikan pada Tabel 3.1.

Tabel 3.1 Konfigurasi hyperparameter utama *fine-tuning*

<i>Hyperparameter</i>	<i>Nilai</i>
<i>Batch size</i>	64
<i>Iterations</i>	2000
<i>Validation interval</i>	200 <i>iterations</i>
<i>Optimizer</i>	Adam
<i>Learning rate</i>	0.0001
Beta1	0.9
<i>Gradient clipping</i>	5
<i>Image size</i>	32 x 100
<i>Input chanel</i>	1 (<i>grayscale</i>)

Hyperparameter pada Tabel 3.1 disusun untuk menjaga stabilitas optimasi dan konsistensi masukan model. Nilai tersebut mengikuti konfigurasi *baseline* yang dilaporkan pada studi komparasi model pengenalan teks (J. Baek et al., 2019). *Batch size* 64 dipilih agar pembaruan bobot tetap stabil dan efisien secara komputasi. Proses optimasi menggunakan Adam dengan *learning rate* 0,0001 dan beta1 0,9 agar konvergensi berjalan halus, serta *gradient clipping* 5 untuk mencegah gradien berlebih. Dari sisi format data, *image size* 32 x 100 dan *input channel* 1 (*grayscale*) digunakan agar input sesuai dengan konfigurasi CRNN. Jumlah *iterations* ditetapkan sebanyak 2000 sebagai batas maksimum pelatihan, bukan sebagai dasar pemilihan model akhir. Nilai ini dipilih agar proses *fine-tuning* memiliki rentang yang cukup untuk menunjukkan perubahan performa dari tahap awal hingga model mulai konvergen. Dengan jumlah tersebut, perilaku pada metrik validasi dapat diamati secara berkala sehingga titik terbaik tidak hanya ditentukan dari iterasi terakhir.

Kemudian, proses pelatihan dijalankan dengan *Iterations* yang ditetapkan yaitu sebanyak 2000. Pada setiap iterasi dilakukan *forward propagation*, perhitungan *loss*, *backpropagation*, dan *optimizer step*. Evaluasi pada set validasi dilakukan secara berkala sesuai *validation interval*, yaitu setiap 200 iterasi serta pada iterasi awal. Hal tersebut dilakukan untuk memantau kemampuan generalisasi model. Selama pelatihan, metrik yang dipantau meliputi *training loss*, *validation loss*, *Word Accuracy*, dan NED. Jika terjadi perbaikan, *checkpoint* terbaik disimpan sebagai *best_accuracy.pth* untuk akurasi terbaik dan *best_norm_ED.pth* untuk NED terbaik. *Checkpoint* disebut terbaik karena dipilih berdasarkan nilai validasi tertinggi yang pernah

dicapai selama pelatihan. Model *best_accuracy.pth* disimpan ketika nilai *Word Accuracy* validasi pada iterasi saat ini lebih tinggi daripada seluruh evaluasi sebelumnya, sedangkan model *best_norm_ED.pth* disimpan ketika nilai NED validasi saat ini lebih tinggi daripada seluruh evaluasi sebelumnya. Jika pada evaluasi berikutnya tidak ada peningkatan, *checkpoint* tidak diperbarui. Proses ini berlanjut hingga *iterations* tercapai, dan model akhir dipilih dari *checkpoint* terbaik sesuai metrik evaluasi yang digunakan.

3.1.5 Evaluasi

Evaluasi dilakukan pada data uji dari pembagian *dataset* utama. Tahap ini bertujuan untuk menilai efektivitas *fine-tuning* terhadap peningkatan kinerja pengenalan teks pada domain label bahan makanan. Pengujian dilakukan dengan membandingkan model pra-latih dan model hasil *fine-tuning* menggunakan metrik berikut:

- a. *Word Accuracy*, proporsi kata yang dikenali benar terhadap total kata pada *ground truth*.
- b. *Normalized Edit Distance* (NED), tingkat kedekatan teks prediksi terhadap *ground truth* berbasis edit distance ternormalisasi.
- c. *Character Error Rate* (CER), tingkat kesalahan prediksi pada level karakter.
- d. *Word Error Rate* (WER), tingkat kesalahan prediksi pada level kata.

Pemilihan metrik tersebut ditujukan untuk menangkap performa sistem dari sisi ketepatan kata, kedekatan string, serta kesalahan karakter dan kata.

Pemilihan metrik tersebut ditujukan untuk menangkap performa sistem dari sisi ketepatan kata, kedekatan string, serta kesalahan karakter dan kata. Dalam penelitian ini, *Word Accuracy* ditetapkan sebagai metrik prioritas penentu karena tujuan utama sistem adalah menghasilkan kata yang tepat secara utuh pada keluaran OCR. Oleh karena itu, pemilihan model utama mengacu pada *checkpoint* dengan nilai *Word Accuracy* validasi tertinggi, sedangkan NED, CER, dan WER digunakan sebagai metrik pendukung untuk analisis kualitas kesalahan secara lebih rinci.

3.1.6 Pembuatan Antarmuka

Antarmuka pengguna akan dibangun menggunakan Streamlit untuk mendukung demonstrasi proses inferensi OCR. Rancangan antarmuka disusun dalam tiga komponen utama, yaitu area pengaturan inferensi, area penyuntingan citra, dan area penayangan hasil. Area pengaturan inferensi digunakan untuk mengatur model pengenalan, model CRAFT, folder input, folder output crop, folder hasil deteksi, serta berkas keluaran gabungan. Area

penyuntingan citra digunakan untuk penyesuaian data masukan sebelum inferensi, mencakup pemilihan citra, rotasi, dan *crop*. Area penayangan hasil digunakan untuk menampilkan informasi status eksekusi, keluaran pembacaan teks, galeri citra yang diproses, serta tabel hasil pengenalan per sub-citra.

- a. Pengguna menyiapkan konfigurasi inferensi sesuai skenario uji.
- b. Pengguna dapat melakukan penyuntingan citra secara opsional sebelum inferensi dijalankan.
- c. Sistem menjalankan inferensi OCR, kemudian antarmuka menampilkan informasi proses dan log eksekusi.
- d. Antarmuka menampilkan keluaran hasil pembacaan teks untuk ditinjau pengguna.

3.1.7 Simulasi Sistem OCR

Simulasi dilakukan pada citra eksternal yang tidak termasuk dalam data latih, validasi, maupun data uji internal. Citra tersebut terdiri dari lima citra label bahan makanan berbahasa Inggris yang diperoleh dari kemasan produk di minimarket, serta tiga citra label obat berbahasa Indonesia yang diperoleh dari sumber daring. Simulasi ini bertujuan untuk mengevaluasi kemampuan sistem dalam kondisi penggunaan nyata dengan variasi tata letak, kualitas citra, serta kompleksitas susunan teks. Selain itu, pengujian pada citra label obat dilakukan untuk melihat kemampuan generalisasi sistem OCR dalam mengenali teks di luar domain label makanan.

Pada pelaksanaannya, setiap citra disiapkan terlebih dahulu melalui penyesuaian area baca dan orientasi teks untuk memastikan teks berada dalam posisi yang optimal sebelum proses inferensi dilakukan. Selanjutnya, sistem memproses citra melalui tahapan deteksi area teks menggunakan CRAFT, pengenalan teks menggunakan CRNN-TRBA, serta rekonstruksi hasil menjadi keluaran teks yang utuh.

Hasil simulasi kemudian digunakan untuk menganalisis keterbacaan teks, mengidentifikasi kesalahan pengenalan karakter, serta membandingkan performa sistem pada domain label makanan dan non-makanan. Analisis ini menjadi dasar untuk menilai apakah sistem yang dikembangkan cenderung berperan sebagai OCR umum atau memiliki kecenderungan spesifik pada domain tertentu.

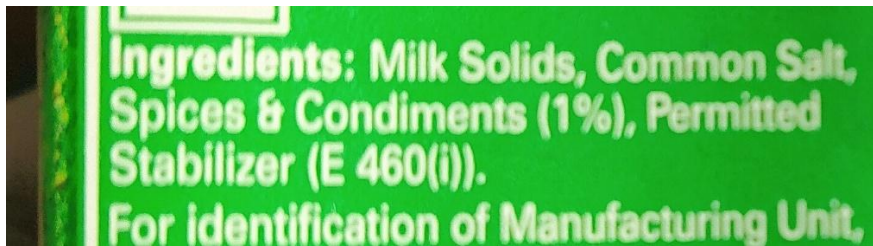
BAB IV

HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

4.1.1 Data Utama Penelitian

Data utama diperoleh dari *dataset* Ingredients Image from Food Label pada platform Kaggle. *Dataset* yang berhasil dikumpulkan berjumlah 182 citra label bahan makanan berbahasa Inggris dengan format citra berwarna dan resolusi yang bervariasi. Contoh variasi citra pada data utama dapat dilihat pada Gambar 4.1, yang menunjukkan label makanan berwarna hijau dengan teks putih. Pada kondisi awal, *dataset* tidak dilengkapi anotasi teks tingkat kata, sehingga dilakukan proses pembentukan pasangan citra-teks untuk kebutuhan pelatihan model pengenalan teks.



Gambar 4.1 Contoh citra *dataset* utama label bahan makanan.

Dari 182 citra tersebut, dilakukan ekstraksi sub-citra berbasis kata menggunakan alur deteksi teks CRAFT dan diperoleh 6812 sub-citra. Selanjutnya, proses pelabelan dilakukan dengan bantuan model CRNN untuk menghasilkan prediksi awal teks, kemudian diverifikasi dan diperbaiki secara manual. Pada tahap verifikasi, sub-citra yang tidak terbaca atau tidak memiliki keterkaitan label yang valid dihapus dari *dataset*. Setelah kurasi selesai, jumlah data yang layak digunakan menjadi 6081 sub-citra, sehingga terdapat penghapusan sebanyak 731 sub-citra atau sekitar 10,73% dari data awal. Dari 6081 sub-citra tersebut, ditemukan 1.986 sub-citra yang dapat dikenali oleh CRNN, tetapi menghasilkan prediksi label yang tidak akurat, sehingga dilakukan proses koreksi manual. Hasil ekstraksi kata dan pasangan labelnya ditampilkan pada Gambar 4.2 dan Gambar 4.3. Gambar 4.2 merupakan hasil ekstraksi citra dari Gambar 4.1, sedangkan Gambar 4.3 merupakan pasangan label sub-citra dari Gambar 4.2.



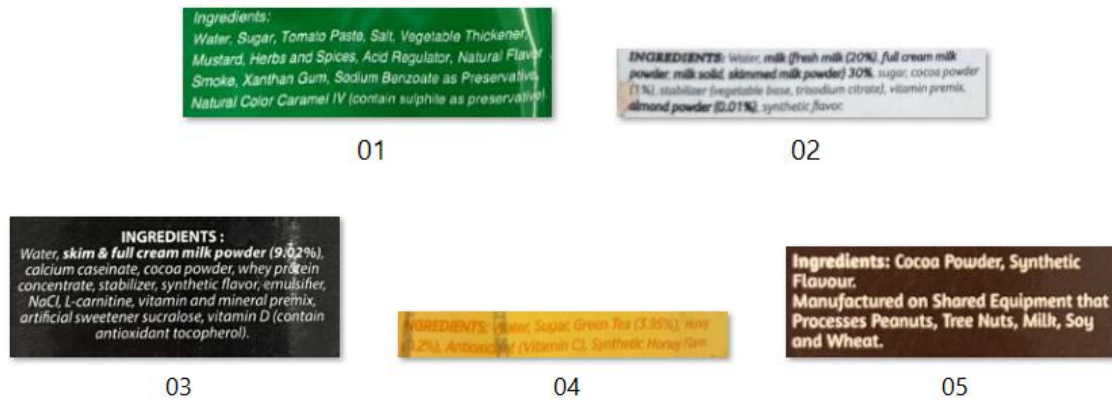
Gambar 4.2 Hasil ekstraksi sub-citra berbasis kata pada contoh citra.

```
word_dataset\in2_box_1.png Ingredients:
word_dataset\in2_box_2.png Milk
word_dataset\in2_box_3.png Solids,
word_dataset\in2_box_4.png Salt,
word_dataset\in2_box_5.png Common
word_dataset\in2_box_6.png Spices
word_dataset\in2_box_7.png Condiments
word_dataset\in2_box_8.png Permitted
word_dataset\in2_box_9.png &
word_dataset\in2_box_10.png %),
word_dataset\in2_box_11.png (1
word_dataset\in2_box_12.png Stabilizer
word_dataset\in2_box_13.png (E
word_dataset\in2_box_14.png 460(i)).
word_dataset\in2_box_15.png Unit,
word_dataset\in2_box_16.png For
word_dataset\in2_box_17.png identification
word_dataset\in2_box_18.png Manufacturing
word_dataset\in2_box_19.png of
```

Gambar 4.3 Pasangan label dan sub-citra setelah verifikasi.

4.1.2 Data Eksternal Simulasi

Selain data utama, penelitian juga menggunakan data eksternal berupa lima citra label bahan makanan berbahasa Inggris yang diambil langsung dari produk kemasan di suatu minimarket yang ditunjukkan pada Gambar 4.4. Selain itu ada juga tiga citra label obat berbahasa Indonesia dari internet yang ditunjukkan pada Gambar 4.5. Data eksternal tersebut akan digunakan pada tahap simulasi sistem OCR.



Gambar 4.4 Lima citra eksternal pada produk kemasan.



Gambar 4.5 Tiga citra eksternal label obat.

Gambar 4.4 digunakan untuk menguji sistem pada citra label makanan yang tidak termasuk dalam data pelatihan. Citra tersebut memiliki variasi warna kemasan, ukuran huruf, dan kualitas pengambilan gambar yang berbeda dari data utama. Sementara itu, Gambar 4.5 digunakan sebagai data simulasi tambahan dari domain label obat. Perbedaan bahasa, struktur informasi, dan istilah pada label obat digunakan untuk melihat kemampuan sistem ketika diterapkan pada data di luar domain utama penelitian.

4.2 Preprocessing

Tahap *preprocessing* dimulai dengan membagi 6081 sub-citra data utama ke dalam data latih, validasi, dan uji dengan target rasio 80:10:10. Hasil pembagian akhir adalah 4868 data latih, 610 data validasi, dan 603 data uji. Proporsi tidak sepenuhnya presisi terhadap rasio karena seluruh sub-citra dari citra induk yang sama ditempatkan pada subset yang sama untuk mencegah kebocoran data. Selanjutnya, berkas anotasi pada setiap subset dikonversi ke format

LMDB agar kompatibel dengan *pipeline* pelatihan model CRNN. Hasil konversi ini menjadi masukan langsung pada tahap *fine-tuning*.

4.3 Pembuatan Sistem OCR

4.3.1 Implementasi Deteksi Teks

Tahap deteksi dimulai dengan memuat bobot model CRAFT pra-latih, menyiapkan direktori keluaran, lalu membaca seluruh citra masukan secara rekursif dari folder input. Setiap citra dipraproses melalui penyesuaian ukuran berbasis rasio aspek dan normalisasi intensitas sebelum inferensi CRAFT dijalankan. Sebagai titik awal proses deteksi, contoh citra masukan yang diproses sistem ditunjukkan pada Gambar 4.6.

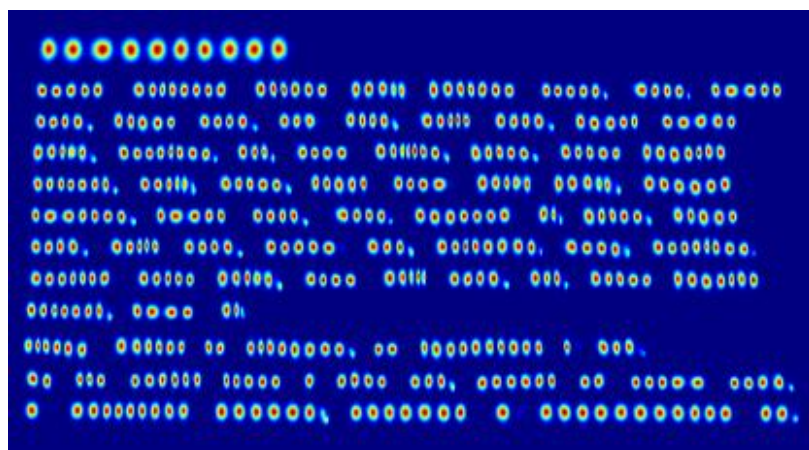


Gambar 4.6 Contoh citra yang akan dideteksi.

Hasil inferensi menghasilkan dua peta utama, yaitu peta region teks dan peta keterhubungan antarkarakter yang ditunjukkan pada Gambar 4.7 dan Gambar 4.8. Peta region teks menunjukkan area pada citra yang memiliki kemungkinan tinggi sebagai teks, sedangkan peta keterhubungan karakter menunjukkan keterhubungan antarkarakter dalam satu kata. Kombinasi kedua peta ini digunakan agar sistem tidak hanya mengenali keberadaan karakter, tetapi juga dapat mengelompokkan karakter menjadi kata yang utuh.



Gambar 4.7 Hasil peta region teks pada citra deteksi.



Gambar 4.8 Hasil peta keterhubungan antarkarakter pada citra deteksi.

Kedua peta tersebut kemudian disimpan sebagai keluaran visual per citra untuk memudahkan inspeksi hasil deteksi. Pengelompokan urutan baca kemudian dilakukan berdasarkan hasil deteksi CRAFT agar susunan baris tetap stabil. Sistem juga menghasilkan berkas *overlay* deteksi saat opsi penyimpanan visualisasi diaktifkan yang ditunjukkan pada Gambar 4.9. *Overlay* ini digunakan untuk memeriksa apakah kotak deteksi sudah mencakup area teks yang benar. Dari keluaran tersebut, tahap deteksi dapat dinilai berhasil apabila kotak berada pada kata yang relevan dan tidak terlalu banyak mencakup area non-teks. Hasil *overlay* juga menjadi dasar untuk tahap pemotongan kata sebelum pengenalan teks dilakukan.



Gambar 4.9 Hasil *overlay* citra deteksi.

4.3.2 Implementasi Pengenalan Teks

Setelah area teks terdeteksi, sistem mengelompokkan kotak deteksi ke dalam baris menggunakan pendekatan *baseline*, kemudian mengurutkan kata dari kiri ke kanan pada setiap baris. Untuk setiap kata, pemotongan citra dilakukan dengan transformasi perspektif agar teks miring tetap terbaca. Jika transformasi perspektif gagal, sistem menggunakan pemotongan cadangan berbasis *axis-aligned box* (kotak sumbu).

Selanjutnya, model CRNN-TRBA dimuat. Sistem kemudian menjalankan pengenalan secara *batch* terhadap seluruh sub-citra kata yang sudah disimpan. Keluaran tahap ini berupa pasangan path sub-citra dan hasil teks prediksi.

4.3.3 Rekonstruksi Teks Label

Hasil pengenalan pada level kata ditunjukkan pada Gambar 4.10, yang menunjukkan bahwa keluaran awal pengenalan berada pada level kata. Setiap baris memuat lokasi potongan citra dan prediksi teks yang dihasilkan model CRNN-TRBA.

outputs\crops\bahan_09_line1_word1.png	INGREDIENTS
outputs\crops\bahan_09_line2_word1.png	Cooked
outputs\crops\bahan_09_line2_word2.png	Marinated
outputs\crops\bahan_09_line2_word3.png	Chicken
outputs\crops\bahan_09_line2_word4.png	(35%)
outputs\crops\bahan_09_line2_word5.png	[Chicken
outputs\crops\bahan_09_line2_word6.png	Breast,
outputs\crops\bahan_09_line2_word7.png	Water,
outputs\crops\bahan_09_line2_word8.png	Tomato
outputs\crops\bahan_09_line3_word1.png	Purée,
outputs\crops\bahan_09_line3_word2.png	Ginger
outputs\crops\bahan_09_line3_word3.png	Purée,
outputs\crops\bahan_09_line3_word4.png	Palm
outputs\crops\bahan_09_line3_word5.png	Olein,

outputs\crops\bahan_09_line3_word6.png Garlic
 outputs\crops\bahan_09_line3_word7.png Puree,
 outputs\crops\bahan_09_line3_word8.png Yogurt
 outputs\crops\bahan_09_line3_word9.png Powder
 outputs\crops\bahan_09_line4_word1.png (Milk),
 outputs\crops\bahan_09_line4_word2.png Cornflour,
 outputs\crops\bahan_09_line4_word3.png Salt,
 outputs\crops\bahan_09_line4_word4.png Green
 outputs\crops\bahan_09_line4_word5.png Chillies,
 outputs\crops\bahan_09_line4_word6.png Spices,
 outputs\crops\bahan_09_line4_word7.png Colour
 outputs\crops\bahan_09_line4_word8.png (Paprika
 outputs\crops\bahan_09_line5_word1.png Extract),
 outputs\crops\bahan_09_line5_word2.png Basil],
 outputs\crops\bahan_09_line5_word3.png Onions,
 outputs\crops\bahan_09_line5_word4.png Single
 outputs\crops\bahan_09_line5_word5.png Cream
 outputs\crops\bahan_09_line5_word6.png (Milk)
 outputs\crops\bahan_09_line5_word7.png (20%),
 outputs\crops\bahan_09_line5_word8.png Chopped
 outputs\crops\bahan_09_line6_word1.png Tomatoes,
 outputs\crops\bahan_09_line6_word2.png Tomato
 outputs\crops\bahan_09_line6_word3.png Paste,
 outputs\crops\bahan_09_line6_word4.png Water,
 outputs\crops\bahan_09_line6_word5.png Rapeseed
 outputs\crops\bahan_09_line6_word6.png Oil,
 outputs\crops\bahan_09_line6_word7.png Spices,
 outputs\crops\bahan_09_line6_word8.png Ginger
 outputs\crops\bahan_09_line7_word1.png Puree,
 outputs\crops\bahan_09_line7_word2.png Garlic
 outputs\crops\bahan_09_line7_word3.png Puree,
 outputs\crops\bahan_09_line7_word4.png Cashew
 outputs\crops\bahan_09_line7_word5.png Nuts,
 outputs\crops\bahan_09_line7_word6.png Coriander,
 outputs\crops\bahan_09_line7_word7.png Honey,
 outputs\crops\bahan_09_line7_word8.png Cornflour,
 outputs\crops\bahan_09_line8_word1.png Unsalted
 outputs\crops\bahan_09_line8_word2.png Butter
 outputs\crops\bahan_09_line8_word3.png (Milk),
 outputs\crops\bahan_09_line8_word4.png Green
 outputs\crops\bahan_09_line8_word5.png Chilli
 outputs\crops\bahan_09_line8_word6.png Puree,
 outputs\crops\bahan_09_line8_word7.png Salt,
 outputs\crops\bahan_09_line8_word8.png Colour
 outputs\crops\bahan_09_line8_word9.png (Paprika
 outputs\crops\bahan_09_line9_word1.png Extract),
 outputs\crops\bahan_09_line9_word2.png Lemon
 outputs\crops\bahan_09_line9_word3.png Oil.
 outputs\crops\bahan_09_line10_word1.png Allergy
 outputs\crops\bahan_09_line10_word2.png Advice!
 outputs\crops\bahan_09_line10_word3.png for
 outputs\crops\bahan_09_line10_word4.png allergens,
 outputs\crops\bahan_09_line10_word5.png see
 outputs\crops\bahan_09_line10_word6.png ingredients
 outputs\crops\bahan_09_line10_word7.png in
 outputs\crops\bahan_09_line10_word8.png bold.
 outputs\crops\bahan_09_line11_word1.png May
 outputs\crops\bahan_09_line11_word2.png also
 outputs\crops\bahan_09_line11_word3.png contain
 outputs\crops\bahan_09_line11_word4.png traces

```

outputs\crops\bahan_09_line11_word5.png of
outputs\crops\bahan_09_line11_word6.png other
outputs\crops\bahan_09_line11_word7.png nuts,
outputs\crops\bahan_09_line11_word8.png peanuts
outputs\crops\bahan_09_line11_word9.png and
outputs\crops\bahan_09_line11_word10.png  sesame
outputs\crops\bahan_09_line11_word11.png  seeds.
outputs\crops\bahan_09_line12_word1.png NO
outputs\crops\bahan_09_line12_word2.png ARTIFICIAL
outputs\crops\bahan_09_line12_word3.png COLOURS,
outputs\crops\bahan_09_line12_word4.png FLAVOURS
outputs\crops\bahan_09_line12_word5.png OR
outputs\crops\bahan_09_line12_word6.png HYDROGENATED
outputs\crops\bahan_09_line12_word7.png FAT.

```

Gambar 4.10 Hasil pengenalan pada level kata.

Hasil tersebut belum langsung membentuk daftar bahan yang utuh. Oleh karena itu, sistem melakukan rekonstruksi dengan membaca indeks urutan kata dari nama berkas, lalu menyusun kembali kata per baris dan per citra. Proses ini menghasilkan tiga jenis keluaran, yaitu gabungan per baris, gabungan per citra, dan gabungan lintas citra dalam satu berkas ringkasan yang ditunjukkan pada Gambar 4.11, Gambar 4.12, dan Gambar 4.13. Melalui rekonstruksi tersebut, keluaran sistem tidak hanya berupa kata terpisah, tetapi menjadi teks daftar bahan yang lebih terstruktur sesuai urutan baca pada label.

Gambar 4.11 memperlihatkan hasil rekonstruksi pada satu baris teks. Pada tahap ini, beberapa prediksi kata dari Gambar 4.10 digabungkan berdasarkan indeks baris yang sama. Kemudian baris-baris tersebut digabungkan menjadi satu per citranya seperti pada Gambar 4.12. Gabungan per citra tersebut kemudian digabung lagi menjadi satu seperti pada Gambar 4.13 untuk memudahkan pemeriksaan banyak citra secara berurutan.

```

Cooked Marinated Chicken (35%) [Chicken Breast, Water, Tomato

```

Gambar 4.11 Gabungan per baris untuk baris ke-2 pada citra deteksi.

INGREDIENTS

Cooked Marinated Chicken (35%) [Chicken Breast, Water, Tomato Puree, Ginger Puree, Palm Olein, Garlic Puree, Yogurt Powder (Milk), Cornflour, Salt, Green Chillies, Spices, Colour (Paprika Extract), Basil], Onions, Single Cream (Milk) (20%), Chopped Tomatoes, Tomato Paste, Water, Rapeseed Oil, Spices, Ginger Puree, Garlic Puree, Cashew Nuts, Coriander, Honey, Cornflour, Unsalted Butter (Milk), Green Chilli Puree, Salt, Colour (Paprika Extract), Lemon Oil.

Allergy Advice! for allergens, see ingredients in bold.

May also contain traces of other nuts, peanuts and sesame seeds.

NO ARTIFICIAL COLOURS, FLAVOURS OR HYDROGENATED FAT.

Gambar 4.12 Gabungan per citra pada citra deteksi.

bahan_08 LOW FAT VANILLA FLAVOURED YOGHURT

INGREDIENTS: Skim Milk, Concentrated Skim Milk, Water, Sugar, Cream (From Milk), Thickeners (1422 (From Maize), 1442 (From Maize)), Milk Solids, Gelatine, Flavours, Acidity Regulators (331, 332, 270, 330), Enzyme (Lactase), Live Cultures. Contains Milk and Milk Products.

bahan_09 INGREDIENTS

Cooked Marinated Chicken (35%) [Chicken Breast, Water, Tomato Puree, Ginger Puree, Palm Olein, Garlic Puree, Yogurt Powder (Milk), Cornflour, Salt, Green Chillies, Spices, Colour (Paprika Extract), Basil], Onions, Single Cream (Milk) (20%), Chopped Tomatoes, Tomato Paste, Water, Rapeseed Oil, Spices, Ginger Puree, Garlic Puree, Cashew Nuts, Coriander, Honey, Cornflour, Unsalted Butter (Milk), Green Chilli Puree, Salt, Colour (Paprika Extract), Lemon Oil.

Allergy Advice! for allergens, see ingredients in bold.

May also contain traces of other nuts, peanuts and sesame seeds.

NO ARTIFICIAL COLOURS, FLAVOURS OR HYDROGENATED FAT.

bahan_10 INGREDIENTS:

REFINED WHEATFLOUR (38%), SUGAR, EDIBLE VEGETABLE OIL (PALM), ROLLED OATS (14%), ALMOND (5%), WHEAT BRAN (2.7%), LIQUID GLUCOSE, RAISING AGENTS [500(ii) 503(ii), 450(i)] MALTODEXTRIN, OAT, FIBRE 6%, INVERT SYRUP, MILK SOLIDS (0.9%), EMULSIFIERS (322.47", 442e) AND IODISED SALT.

CONTAINS PERMITTED NATURAL FOOD COLOUR (150d) AND ADDED FLAVOURS [NATURE IDENTICAL ARTIFICIAL FLAVOURING SUBSTANCES (BUTTER, MILK)]

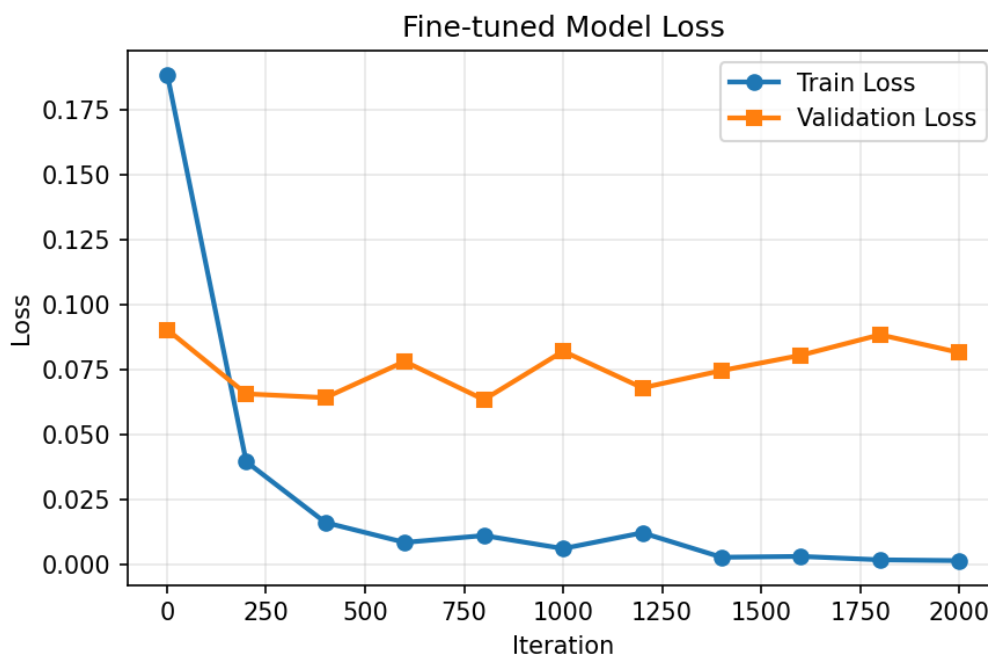
(Numbers in brackets as per International Numbering System)

Gambar 4.13 Contoh gabungan lintas 3 citra.

4.4 Fine-tuning Model Pengenalan Teks

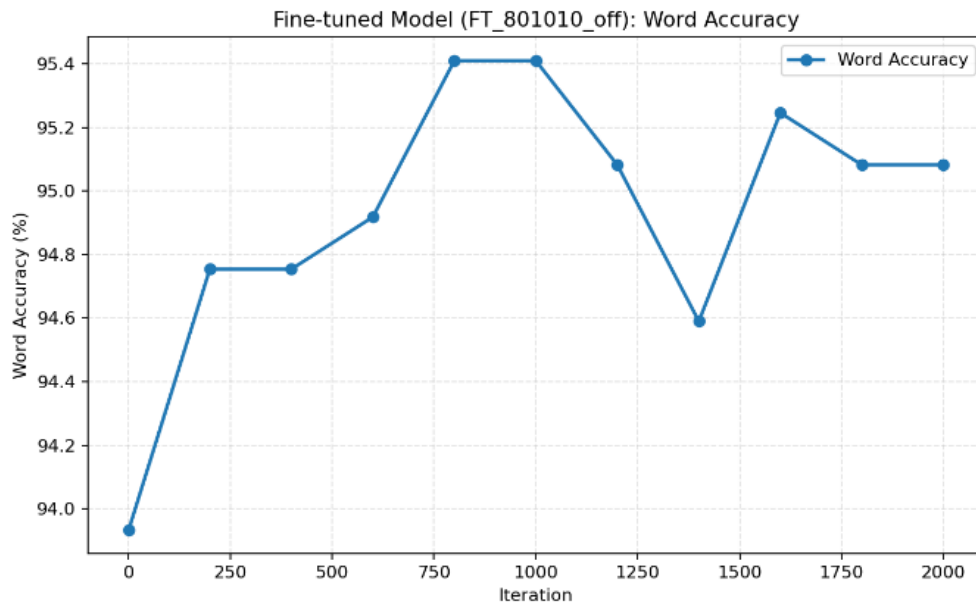
4.4.1 Hasil Pelatihan *Fine-tuning*

Fine-tuning dilakukan pada model CRNN pra-latih dengan konfigurasi pelatihan yang telah ditetapkan sebelumnya, sedangkan model CRAFT tetap digunakan sebagai model deteksi pra-latih tanpa pelatihan ulang. Pelatihan dijalankan sampai 2000 iterasi. Pada setiap iterasi, model menjalankan *forward propagation*, perhitungan *loss*, *backpropagation*, dan *optimizer step*. Validasi dilakukan secara berkala setiap 200 iterasi serta pada iterasi awal. Perbandingan kurva *Training loss* dan *Validation loss* selama *fine-tuning* dapat dilihat pada Gambar 4.14.

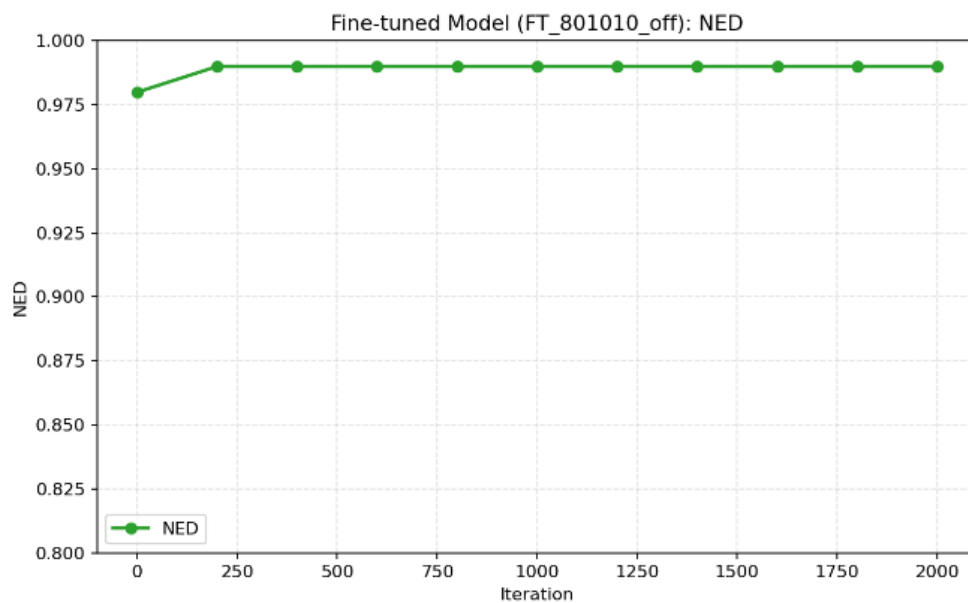


Gambar 4.14 Kurva *training loss* vs *validation loss* saat *fine-tuning*

Berdasarkan log pelatihan, *Training Loss* menurun konsisten dari 0,18861 pada iterasi awal menjadi 0,00132 pada iterasi ke-2000. Pada sisi validasi, *Validation Loss* mencapai nilai terendah 0,06339 pada iterasi ke-800, kemudian cenderung berfluktuasi pada iterasi berikutnya. Nilai *Word Accuracy* meningkat dari 93,934% pada iterasi awal, mencapai nilai tertinggi 95,410% pada rentang iterasi 800-1000, lalu berada di 95,082% pada iterasi akhir. Selain itu, NED meningkat dari 0,98 pada validasi awal menjadi 0,99 sejak iterasi 200 dan bertahan pada kisaran tersebut hingga akhir pelatihan. Pola perubahan *Word Accuracy* dan NED selama validasi divisualisasikan pada Gambar 4.15 dan Gambar 4.16.



Gambar 4.15 Kurva *Word Accuracy* saat validasi *fine-tuning*



Gambar 4.16 Kurva *NED* saat validasi *fine-tuning*

4.4.2 Pembahasan Hasil *Fine-tuning*

Hasil pelatihan menunjukkan bahwa model berhasil beradaptasi terhadap domain teks daftar bahan makanan, terutama pada fase awal hingga menengah pelatihan. Penurunan *training loss* yang stabil menandakan optimasi berjalan efektif, sedangkan pola *validation loss* yang kembali naik setelah titik terbaik menunjukkan mulai muncul kecenderungan *overfitting*

ringan, yaitu kondisi ketika model terus membaik pada data latih, tetapi kemampuan generalisasinya pada data validasi tidak lagi meningkat secara konsisten.

Implikasi dari kondisi tersebut adalah checkpoint akhir pada iterasi ke-2000 tidak langsung dipilih sebagai model utama, meskipun memiliki *training loss* paling rendah. Pemilihan model akhir perlu mempertimbangkan performa validasi karena model akan digunakan pada citra baru yang tidak dilibatkan dalam pelatihan. Jika hanya mengacu pada *training loss*, model berisiko terlalu menyesuaikan diri terhadap data latih dan menghasilkan performa yang kurang stabil pada data uji.

Berdasarkan pertimbangan tersebut, checkpoint pada sekitar iterasi ke-800 dipilih sebagai model yang paling representatif. Pada titik ini, *validation loss* mencapai nilai minimum sebesar 0,06339, *Word Accuracy* validasi mencapai 95,410%, dan NED berada pada 0,99. Kombinasi tersebut menunjukkan bahwa model tidak hanya memiliki kesalahan validasi paling rendah, tetapi juga menghasilkan akurasi kata dan kedekatan prediksi yang tinggi terhadap teks acuan. Dengan demikian, checkpoint iterasi ke-800 dipilih karena memberikan keseimbangan terbaik antara kemampuan mempelajari data latih dan kemampuan generalisasi pada data validasi.

4.5 Evaluasi

4.5.1 Metrik Evaluasi

Evaluasi model menggunakan empat metrik utama, yaitu *Word Accuracy*, NED, CER, dan WER.

4.5.2 Hasil Perbandingan Model

Perbandingan performa dilakukan pada tiga model, yaitu model pra-latih, model hasil *fine-tuning* dengan akurasi validasi terbaik, model hasil *fine-tuning* dengan NED validasi terbaik, dan Tesseract. Hasil pengujian ditunjukkan pada Tabel 4.1.

Tabel 4.1 Hasil evaluasi model pada data uji.

Model	<i>Word Accuracy</i>	NED	CER	WER
Pra-latih	90.381%	0.946	0.066	0.096
Akurasi terbaik	94.693%	0.987	0.015	0.053
NED terbaik	94.030%	0.985	0.017	0.060
Tesseract	70.813%	0.894	0.089	0.300

Hasil evaluasi menunjukkan bahwa kedua model hasil *fine-tuning* memberikan peningkatan performa dibandingkan model pra-latih pada seluruh metrik utama. Model dengan akurasi terbaik memberikan nilai *Word Accuracy* dan NED tertinggi, sekaligus CER dan WER terendah. Model dengan NED terbaik juga menunjukkan performa yang lebih baik dibandingkan model pra-latih, namun sedikit berada di bawah model akurasi terbaik.

Secara kuantitatif, *fine-tuning* memberikan peningkatan yang jelas dibandingkan model pra-latih. Jika dibandingkan antara model pra-latih dan model akurasi terbaik, akurasi kata meningkat dari 90,381% menjadi 94,693%, NED meningkat dari 0,946 menjadi 0,987, CER turun dari 0,066 menjadi 0,015, dan WER turun dari 0,096 menjadi 0,053. Hasil tersebut menunjukkan bahwa model semakin sesuai dengan karakteristik teks daftar bahan makanan dan kesalahan prediksi pada tingkat karakter maupun kata berhasil ditekan. Namun, peningkatan ini belum dapat dianggap sempurna. Nilai WER 0,053 masih menunjukkan bahwa sekitar 5 dari 100 kata berpotensi salah dikenali.

Peningkatan performa setelah *fine-tuning* ini berasal dari komponen BiLSTM dan Attention yang tetap dilatih, sedangkan TPS dan ResNet dipertahankan sebagai ekstraktor fitur visual dasar. Dengan skema ini, BiLSTM menjadi lebih efektif memanfaatkan konteks dua arah antarkarakter dalam kata, sementara Attention menjadi lebih tepat dalam melakukan *alignment* fitur urutan karakter saat *decoding*. Akibatnya, model lebih akurat pada prediksi kata utuh dan lebih kecil pada kesalahan edit di tingkat karakter, yang tercermin pada kenaikan *Word Accuracy* dan NED serta penurunan CER dan WER. Dengan demikian, peningkatan yang diperoleh dapat diinterpretasikan sebagai keberhasilan adaptasi *sequence modeling* dan *decoding*.

Pada Tabel 4.1 juga terdapat Tesseract sebagai OCR umum. Tesseract memperoleh *Word Accuracy* sebesar 70,813%, NED sebesar 0,894, CER sebesar 0,089, dan WER sebesar 0,300. Nilai tersebut merupakan performa terendah pada pengujian ini. Bahkan sebelum *fine-tuning* dilakukan, model pra-latih sudah memberikan hasil yang lebih baik daripada Tesseract, yaitu *Word Accuracy* 90,381% dibandingkan 70,813%, NED 0,946 dibandingkan 0,894, CER 0,066 dibandingkan 0,089, dan WER 0,096 dibandingkan 0,300. Selisih tersebut menunjukkan bahwa model pra-latih memiliki kemampuan pengenalan teks yang lebih sesuai untuk citra teks dibandingkan Tesseract pada data uji penelitian ini.

Keunggulan model pra-latih terhadap Tesseract dapat dijelaskan dari karakteristik pendekatannya. Model CRNN-TRBA menggunakan kombinasi TPS, ResNet, BiLSTM, dan Attention yang dirancang untuk mengenali teks pada citra dengan variasi bentuk, orientasi, dan

kualitas visual. TPS membantu menormalkan bentuk teks, ResNet mengekstraksi fitur visual, BiLSTM memanfaatkan konteks urutan karakter, sedangkan Attention membantu proses decoding karakter. Sebaliknya, Tesseract sebagai OCR umum lebih sensitif terhadap kualitas segmentasi, kontras, kemiringan, dan kerapatan teks. Pada label bahan makanan, teks sering berukuran kecil, rapat, melengkung, atau memiliki latar yang tidak seragam, sehingga Tesseract menghasilkan kesalahan kata yang lebih tinggi. Oleh karena itu, hasil evaluasi menunjukkan bahwa model pra-latih CRNN-TRBA sudah lebih sesuai daripada Tesseract, dan *fine-tuning* lebih lanjut membuat performanya meningkat pada domain label bahan makanan.

4.6 Pembuatan Antarmuka

Pada area pengaturan inferensi yang ditunjukkan pada Gambar 4.17, antarmuka menyediakan kolom konfigurasi “Saved model path” untuk menentukan model CRNN-TRBA, “Input folder” untuk lokasi citra masukan, “CRAFT model path” untuk model deteksi teks, “Crops folder” untuk keluaran potongan kata, “Results folder” untuk keluaran hasil deteksi dan pengenalan, serta “Merged output file” untuk berkas gabungan teks hasil OCR. Area ini juga memuat opsi *checkbox* yang memengaruhi perilaku *pipeline*, yaitu “Use CRAFT refiner” untuk mengaktifkan deteksi per baris, “Save overlay images” untuk menyimpan visualisasi kotak deteksi pada citra, “Force sensitive charset” untuk mempertahankan perbedaan huruf besar dan kecil, “Use PAD” untuk menerapkan *padding* saat penyesuaian ukuran input pengenalan, “Use RGB recognizer input” untuk menggunakan masukan tiga kanal warna pada pengenalan teks, dan “Use CUDA” untuk menjalankan inferensi pada GPU jika tersedia. Opsi-opsi *checkbox* tersebut diadaptasi dari konfigurasi bawaan pada kode model CRNN-TRBA sehingga semua fitur pada model dapat diimplementasikan (J. Baek et al., 2019).

Food Ingredient OCR Runner

Run CRAFT + TRBA OCR from UI and inspect merged output quickly.

Saved model path saved_models/FT_801010_off/best_accuracy.	Crops folder outputs/crops
Input folder images/gambar_simulasi	Results folder outputs/craft
CRAFT model path saved_models/craft_mlt_25k.pth	Merged output file outputs/merged.txt
☐ Use CRAFT refiner ☒ Save overlay images ☒ Force sensitive charset	☐ Use PAD ☐ Use RGB recognizer input ☒ Use CUDA
Run OCR	

> Image Editor (Optional)

Gambar 4.17 Tampilan area pengaturan inferensi.

Selanjutnya, eksekusi dapat dijalankan melalui tombol “Run OCR” yang kemudian akan menampilkan perintah yang dieksekusi sebagai jejak konfigurasi percobaan seperti pada Gambar 4.18.

```
C:\Users\rakee\AppData\Local\Programs\Python\Python311\python.exe demo.py --saved_model saved_r
```

Running OCR pipeline...

Gambar 4.18 Tampilan perintah yang dieksekusi.

Sebelum tombol “run OCR” dijalankan, area penyuntingan citra dapat digunakan untuk melakukan rotasi dan pemotongan citra sebagai penyesuaian awal ketika orientasi atau area teks pada gambar belum ideal seperti pada Gambar 4.19. Citra yang ingin di sunting dipilih melalui “Select image to edit”, mengatur rotasi melalui “Rotate angle (deg)”, dan dapat

mengaktifkan “Keep full image after rotation” agar seluruh konten tetap tercakup setelah rotasi. Jika diperlukan pemotongan area baca, pengguna dapat mengaktifkan “Enable crop” yang ditunjukkan pada Gambar 4.20, lalu menentukan batas potong melalui “Crop X”, “Crop Y”, “Crop Width”, dan “Crop Height”. Hasil penyuntingan ditampilkan pada panel “Preview”. Penyimpanan disediakan dalam dua mode, yaitu “Save (overwrite selected image)” untuk menimpa berkas terpilih dan “Save as new file” untuk menyimpan salinan baru.

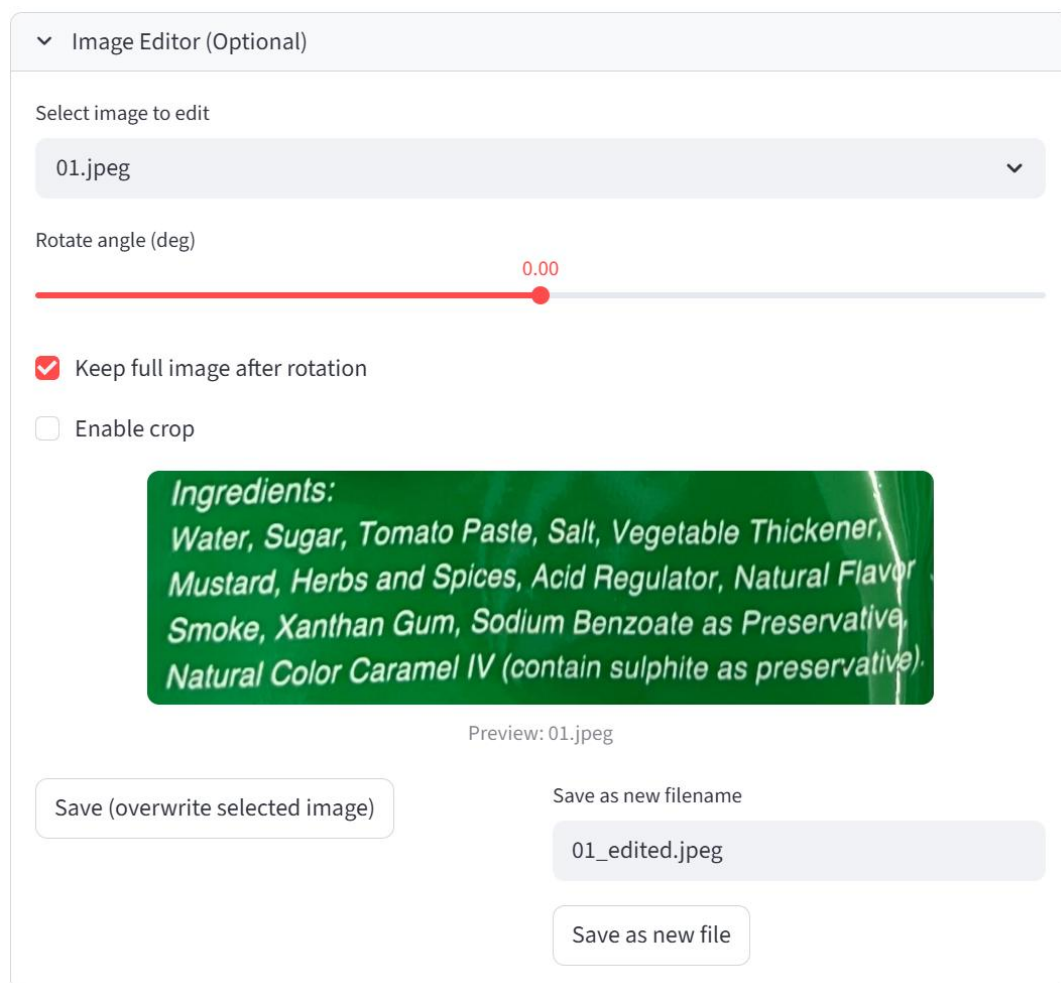


Image Editor (Optional)

Select image to edit

01.jpeg

Rotate angle (deg)

0.00

☒ Keep full image after rotation

☐ Enable crop

Ingredients:
Water, Sugar, Tomato Paste, Salt, Vegetable Thickener,
Mustard, Herbs and Spices, Acid Regulator, Natural Flavor
Smoke, Xanthan Gum, Sodium Benzoate as Preservative,
Natural Color Caramel IV (contain sulphite as preservative).

Preview: 01.jpeg

Save (overwrite selected image)

Save as new filename

01_edited.jpeg

Save as new file

Gambar 4.19 Bagian penyuntingan citra sebelum inferensi.

☒ Enable crop

Crop X
0

Crop Y
0

Crop Width
1280

Crop Height
381

Gambar 4.20 Bagian penyuntingan potongan citra jika *checkbox* “Enable crop” diaktifkan.

Pada area penayangan hasil, tampilan pertama yang disajikan adalah bagian “Stdout” yang ditunjukkan pada Gambar 4.21. Bagian ini digunakan untuk menampilkan status proses dan jejak eksekusi setelah tombol “Run OCR” dijalankan. “Stdout” berisi keluaran proses utama *pipeline* dan jika terjadi kesalahan pada proses *running* maka bagian ini akan menampilkan *stderr* yang menandakan kegagalan pada tahap tertentu. Dengan begini, sumber masalah dapat ditelusuri secara lebih terarah pada level antarmuka.

```

Demo finished successfully.

Stdout

stdout

Recognizer params 32 100 20 1 512 256 96 25 TPS ResNet BiLSTM Attn
Loading recognizer weights from saved_models/FT_801010_off/best_accuracy.pth
Found 5 images in images\gambar_simulasi
images\gambar_simulasi\01.jpeg -> saved 31 crops
images\gambar_simulasi\02.jpeg -> saved 33 crops
images\gambar_simulasi\03.jpeg -> saved 35 crops
images\gambar_simulasi\04.jpeg -> saved 15 crops
images\gambar_simulasi\05.jpeg -> saved 18 crops
Total crops written: 132
Recognition results saved to outputs\craft\recognized.txt

```

Gambar 4.21 Tampilan bagian “Stdout”.

Tampilan kedua adalah bagian “Merged Output” dan “Input Images” yang ditampilkan bersebelahan pada Gambar 4.22. Panel “Merged Output” menampilkan isi *merged.txt* sebagai hasil rekonstruksi teks akhir per citra. Di sebelahnya, panel “Input Images” menampilkan seluruh citra masukan yang diproses untuk memudahkan pencocokan antara teks keluaran dan citra sumber. Format berdampingan ini membantu evaluasi cepat antara citra masukan dan hasil prediksi sistem OCR.

Merged Output

merged.txt

01 Ingredients:

Water, Sugar, Tomato Paste, Salt, Vegetable Thickener, Mustard, Herbs and Spices, Acid Regulator, Natural Flavor
Smoke, Xanthan Gum, Sodium Benzoate as Preservative,
Natural Color Caramel IV (contain sulphite as preservative).

02 INGREDIENTS: Water, milk (fresh milk (20%), full cream milk powder, milk solid, skimmed milk powder) 30%, sugar, cocoa powder (1 %), stabilizer (vegetable base, trisodium citrate), vitamin premix, almond powder (0.01%), synthetic flavor.

03 INGREDIENTS:

Input Images

Water, Sugar, Tomato Paste, Salt, Vegetable Thickener, Mustard, Herbs and Spices, Acid Regulator, Natural Flavor
Smoke, Xanthan Gum, Sodium Benzoate as Preservative, Natural Color Caramel IV (contain sulphite as preservative).

01.jpeg

INGREDIENTS: Water, milk (fresh milk (20%), full cream milk powder, milk solid, skimmed milk powder) 30%, sugar, cocoa powder (1 %), stabilizer (vegetable base, trisodium citrate), vitamin premix, almond powder (0.01%), synthetic flavor.

02.jpeg

INGREDIENTS :
Water, **skim & full cream milk powder (9.02%)**, calcium caseinate, cocoa powder, whey protein concentrate, stabilizer, synthetic flavor, emulsifier, NaCl, L-carnitine, vitamin and mineral premix, artificial sweetener sucralose, vitamin D (contain antioxidant tocopherol).

03.jpeg

Gambar 4.22 Tampilan bagian “Merged Output” dan “Input Images”.

Tampilan ketiga yang ditunjukkan pada Gambar 4.23 adalah bagian “Recognized Table” yang menampilkan rincian hasil pengenalan pada level potongan kata. Tabel ini memuat kolom “crop_path” sebagai lokasi sub-citra hasil pemotongan dan kolom “text” sebagai prediksi teks setiap potongan. Bagian ini digunakan untuk analisis kesalahan yang lebih detail dengan mengidentifikasi potongan mana yang menghasilkan prediksi benar atau salah sebelum rekonstruksi ke teks gabungan.

Recognized Table

	crop_path	text
0	outputs\crops\01_line1_word1.png	Ingredients:
1	outputs\crops\01_line2_word1.png	Water,
2	outputs\crops\01_line2_word2.png	Sugar,
3	outputs\crops\01_line2_word3.png	Tomato
4	outputs\crops\01_line2_word4.png	Paste,
5	outputs\crops\01_line2_word5.png	Salt,
6	outputs\crops\01_line2_word6.png	Vegetable
7	outputs\crops\01_line2_word7.png	Thickener,

Gambar 4.23 Tampilan bagian “Recognized Table”.

4.6.1 Hasil Uji Fungsional Antarmuka

Pengujian fungsional pada Subbab Implementasi Antarmuka menunjukkan bahwa antarmuka dapat menjalankan alur utama dengan baik, mulai dari menerima citra masukan, mengeksekusi proses OCR, hingga menampilkan keluaran hasil pembacaan teks. Setelah inferensi selesai, antarmuka menampilkan informasi proses, hasil gabungan teks, daftar citra yang diproses, dan tabel hasil pengenalan kata.

Secara umum, antarmuka sudah memadai sebagai media demonstrasi sistem karena menyederhanakan proses eksekusi *pipeline* OCR dan mempercepat peninjauan hasil. Kelebihan utama antarmuka terletak pada integrasi proses dari input hingga output dalam satu alur kerja. Namun, masih terdapat beberapa kekurangan yang perlu diperbaiki.

Pada konfigurasi penelitian ini, pada area pengaturan inferensi *checkbox* yang diaktifkan adalah “Save overlay images”, “Force sensitive charset”, dan “Use CUDA”. Sementara itu, “Use CRAFT refiner”, “Use PAD”, dan “Use RGB recognizer input” tidak diaktifkan pada konfigurasi karena permasalahan pada efisiensi proses dan kestabilan akurasi hasil. Pada pengujian opsi “Use CRAFT refiner”, sistem memang menghasilkan area baris terlebih dahulu, tetapi waktu proses meningkat karena alur deteksi berjalan dua tahap, yaitu deteksi awal lalu deteksi lanjutan pada hasil baris. Dampaknya, durasi inferensi dapat mendekati dua kali proses standar. Selain itu, pada beberapa citra ditemukan crop baris yang terpotong di bagian ujung sehingga baris tidak utuh, dan pada kasus lain lebar area baris terlalu besar sehingga teks dari baris di bawah ikut terbaca. Pada pengujian opsi Use PAD, aktivasi *padding* tidak meningkatkan kualitas prediksi pada data penelitian ini, tetapi justru menurunkan ketepatan

hasil pengenalan. Secara teori, *padding* menambah area kosong saat citra dinormalisasi ke ukuran masukan tetap, sehingga proporsi karakter efektif menjadi lebih kecil dan detail karakter halus atau tanda baca lebih mudah tertukar.

Selain itu, perlu dijelaskan juga bahwa antarmuka ini masih berfokus pada kebutuhan demonstrasi dan belum diarahkan sebagai aplikasi utuh. Oleh karena itu, pengembangan lanjutan masih diperlukan, terutama pada aspek validasi masukan, pengelolaan kesalahan, dan peningkatan kenyamanan pengguna untuk skenario penggunaan yang lebih luas.

4.7 Simulasi Sistem OCR

4.7.1 Hasil Simulasi Lima Citra Label Makanan

Simulasi dilakukan pada lima citra label makanan yang kemudian dilakukan evaluasi dengan membandingkan teks hasil prediksi terhadap teks acuan. Hasil evaluasi kuantitatif lima citra label makanan disajikan pada Tabel 4.2.

Tabel 4.2 Hasil evaluasi lima citra label makanan.

Citra Ke-	Jumlah kata prediksi/Total kata	Word Accuracy	NED	CER	WER
1	31/31	100.000%	1.000	0.000	0.000
2	33/32	93.750%	0.954	0.004	0.062
3	35/34	88.235%	0.899	0.011	0.118
4	15/14	78.571%	0.828	0.037	0.214
5	18/18	94.444%	0.944	0.007	0.056
1-5	132/129	92.248%	0.988	0.009	0.078

Berdasarkan tabel tersebut, sistem menghasilkan 132 kata prediksi dari 129 kata acuan. Nilai *Word Accuracy* sebesar 92,248% menunjukkan bahwa sebagian besar kata pada lima citra berhasil dikenali dengan benar. Nilai NED sebesar 0,988 juga menunjukkan bahwa teks prediksi memiliki kedekatan yang tinggi terhadap teks acuan. Selain itu, nilai CER sebesar 0,009 menunjukkan bahwa kesalahan pada tingkat karakter relatif kecil, sedangkan nilai WER sebesar 0,078 menunjukkan bahwa sekitar 7,8% kata masih mengalami kesalahan pada tingkat kata.

Hasil visual dari proses simulasi ditunjukkan pada Gambar 4.24. Pada gambar tersebut, highlight kuning digunakan untuk menandai bagian karakter atau kata yang tidak sesuai dengan teks acuan, sehingga lokasi kesalahan dapat dilihat langsung pada hasil prediksi sistem.

prediksi dari 301 kata acuan, dengan *Word Accuracy* sebesar 85,382%, NED sebesar 0,965, CER sebesar 0,034, dan WER sebesar 0,146. Hasil keseluruhan simulasi pada tiga citra label obat disajikan pada Tabel 4.3.

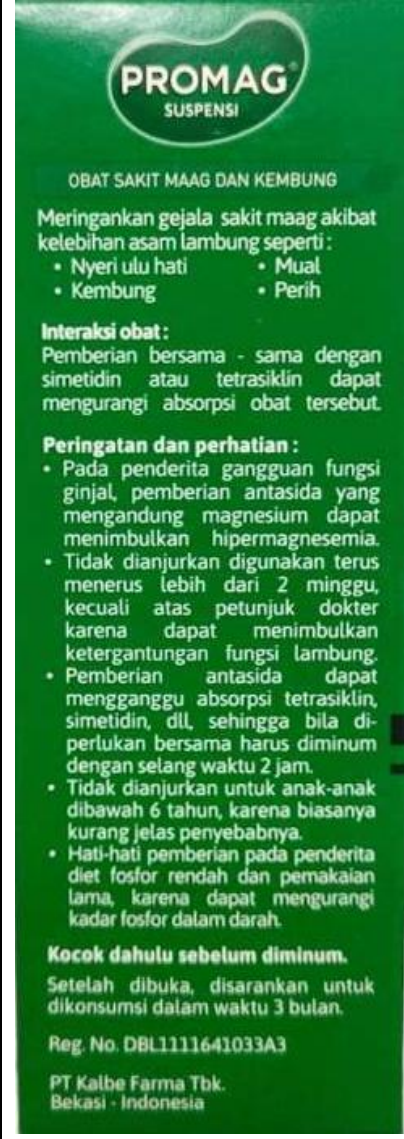
Tabel 4.3 Hasil evaluasi tiga citra label obat.

Citra Ke-	Jumlah kata prediksi/Total kata	<i>Word Accuracy</i>	NED	CER	WER
1	79/78	79.487%	0.933	0.067	0.205
2	140/141	88.652%	0.976	0.024	0.113
3	84/82	85.366%	0.986	0.014	0.146
1-3	303/301	85.382%	0.965	0.034	0.146

Untuk memperjelas bentuk kesalahan yang muncul, hasil kuantitatif dilengkapi dengan perbandingan antara citra masukan dan teks hasil prediksi. Perbandingan tersebut disajikan pada Tabel 4.4, sehingga keluaran sistem dapat ditinjau tidak hanya dari nilai metrik, tetapi juga dari kecocokan teks prediksi terhadap tampilan label obat pada masing-masing citra.

Tabel 4.4 Perbandingan citra dan prediksi.

Citra	Prediksi
 INTUNAL-F[®] TABLET KOMPOSISI: Tiap tablet mengandung: Paracetamol..... 500,00 mg Phenylpropanolamine HCl..... 15,00 mg Dexchlorpheniramine maleate... 2,00 mg Dextromethorphan HBr..... 15,00 mg Guaifenesin..... 50,00 mg CARA KERJA OBAT: Bekerja sebagai analgesik-antipiretik, ekspektoran, anti-histamin, antitusif dan dekongestan hidung. INDIKASI: Untuk meringankan gejala-gejala flu seperti demam, sakit kepala, hidung tersumbat, dan bersin-bersin yang disertai batuk. KONTRAINDIKASI: • Lihat Box Warning • Penderita dengan gangguan jantung dan diabetes mellitus • Penderita dengan gangguan fungsi hati yang berat • Penderita yang hipersensitif terhadap komponen obat ini	INTUNAL-F TABLET KOMPOSISI: Tiap tablet mengandung: Paracetamoi, 500,00 mg Phenyloropropanolamine HCl. 15, 00 mg Dexchlorpheniramine maleate... 2,00 mg Dextromethorphan HBr. 15,00 mg Guaifenesin. 50,00 mg CARA KERJA OBAT: Bekerja sebagai analgesk-antiplatek, ekspektoran, anti-histamin, antitusif dan dekongestan hidung. INDIKASI: Untuk meringankan gejala-gejala flu seperti demam, sakit kepala, hidung tersumbat, dan bersin- bersin yang disertai batuk. KONTRAINDIKASI: Lihat Box Warning Penderita dengan gangguan jantung dan diabetes mellitus Penderita Penderita dengan gangguan fungsi hati yang berat yang hipersensitif terhadap komponen obat ini

 PROMAG SUSPENSIO OBAT SAKIT MAAG DAN KEMBUNG Meringankan gejala sakit maag akibat kelebihan asam lambung seperti : • Nyeri ulu hati • Mual • Kembung • Perih Interaksi obat: Pemberian bersama - sama dengan simetidin atau tetrasiklin dapat mengurangi absorpsi obat tersebut. Peringatan dan perhatian : • Pada penderita gangguan fungsi ginjal, pemberian antasida yang mengandung magnesium dapat menimbulkan hipermagnesemia. • Tidak dianjurkan digunakan terus menerus lebih dari 2 minggu, kecuali atas petunjuk dokter karena dapat menimbulkan ketergantungan fungsi lambung. • Pemberian antasida dapat mengganggu absorpsi tetrasiklin, simetidin, dll, sehingga bila diperlukan bersama harus diminum dengan selang waktu 2 jam. • Tidak dianjurkan untuk anak-anak dibawah 6 tahun, karena biasanya kurang jelas penyebabnya. • Hati-hati pemberian pada penderita diet fosfor rendah dan pemakaian lama, karena dapat mengurangi kadar fosfor dalam darah. Kocok dahulu sebelum diminum. Setelah dibuka, disarankan untuk dikonsumsi dalam waktu 3 bulan. Reg. No. DBL1111641033A3 PT Kalbe Farma Tbk. Bekasi - Indonesia	PROMAG SUSPENSIO) OBAT SAKIT MAAG DAN KEMBUNG Meringankan gejala sakit maag akibat kelebihan asam lambung seperti: Nyeri ulu hati Mual Kembung Perih Interaksi obat: Pemberian bersama sama dengan simetidin atau tetrasiklin dapat mengurangi absorpsi obat tersebut Peringatan dan perhatian Pada penderita gangguan fungsi ginjal, pemberian antasida yang mengandung magnesium dapat menimbulkan hipermagnesemia. Tidak dianjurkan digunakan terus menerus lebih dari 2 minggu, kecuali atas petunjuk dokter karena dapat menimbulkan ketergantungan fungsi lambung. Pemberian antasida dapat mengganggu absorpsi tetrasiklin, simetidin, dll, sehingga bila diperlukan bersama harus diminum dengan waktu jam. Tidak inselants untuk anak anak dibawah 6 tahun, karena biasanya kurang Jelas penyebabnya. Hati-haul pemberian pada penderita diet foster rendah dan pemakaian lama, karena dapat mengurangi kadar foster dalam darah, Kocok dahulu sebelum diminum. Setelah dibuka, disarankan untuk dikonsurssi dalam waktu bulan. Reg. No. DBL111164103343 PT Kalbe Farma Tbk. Bekasi Indonesia
KOMPOSISI Tiap tube 5 mL mengandung : Sodium Lauryl Sulfoacetate0,045 g Sodium Citrate0,450 g Sorbitol4,465 g Polyethylene Glycol 4000,625 g INDIKASI Microlax® obat pencahar untuk mengatasi susah buang air besar (sembelit), khususnya diberikan pada penderita yang harus tinggal di tempat tidur; orang dewasa, orang tua, anak-anak dan wanita hamil. KONTRAINDIKASI Microlax® jangan digunakan pada penderita penyakit wasir akut dan penderita radang usus besar EFEK SAMPING Belum pernah ada laporan adanya efek samping. Penggunaan berlebihan dapat menyebabkan diare dan kekurangan cairan.	KOMPOSISI Tiap tube 5mL mengandung Sodium Lauryl Sulfoacetate 0,045 g Sodium Citrate 0, 450 g Sorbitol. 4, 465 Polyethylene Glycol 400 0, 625 g INDIKASI Microlax* obat pencahar untuk mengatasi susah buang air besar (sembelit), khususnya diberikan pada penderita yang harus tinggal di tempat tidur; orang dewasa, orang tua, anak-anak dan wanita hamil. KONTRAINDIKASI Microlax jangan digunakan pada penderita penyakit wasir akut dan penderita radang usus besar EFEK SAMPING Belum pernah ada laporan adanya efek samping. Penggunaan berlebihan dapat menyebabkan diare dan kekurangan cairan

Berdasarkan tabel tersebut, citra pertama menunjukkan bahwa bagian utama label seperti “INTUNAL-F”, “TABLET”, “KOMPOSISI”, “CARA KERJA OBAT”, “INDIKASI”, dan “KONTRAINDIKASI” dapat terbaca. Namun, beberapa kesalahan masih muncul pada level karakter, misalnya “Paracetamoi” untuk “Paracetamol”, “analgesk-antipletik”, dan “gejala-gejala”. Kesalahan tersebut menunjukkan bahwa kata teknis dan huruf yang bentuknya mirip masih berpengaruh terhadap kestabilan prediksi.

Pada citra kedua, sistem mampu mengenali nama produk “PROMAG” serta beberapa bagian penting seperti “OBAT SAKIT MAAG DAN KEMBUNG” dan “Interaksi obat”. Meskipun demikian, kesalahan karakter masih terlihat pada prediksi seperti “SUSPENS)” untuk “SUSPENS”, “dil” untuk “dll”, “Hati-haul”, “fostor”, dan “dikonsurssi”. Kesalahan ini berkaitan dengan kondisi teks yang kecil, rapat, serta adanya variasi kualitas citra yang dapat memengaruhi proses deteksi dan pengenalan karakter.

Pada citra ketiga, struktur informasi utama seperti “KOMPOSISI”, “INDIKASI”, “KONTRAINDIKASI”, dan “EFEK SAMPING” dapat dikenali, begitu pula beberapa komponen seperti “Sodium Lauryl Sulfoacetate”, “Sodium Citrate”, dan “Polyethylene Glycol 400”. Namun, sebagian prediksi belum sepenuhnya tepat, misalnya “Sorbitol. 4, 465” yang kehilangan satuan ukuran, serta penggunaan simbol pada “Microlax*” yang belum konsisten. Secara keseluruhan, hasil pada Tabel 4.3 dan Tabel 4.4 menunjukkan bahwa sistem masih mampu membaca teks pada label obat, tetapi performanya lebih rendah dibandingkan domain target karena model pengenalan disesuaikan menggunakan data label makanan. Penerapan pada label obat memerlukan data tambahan dan *fine-tuning* khusus agar nilai *Word Accuracy* dan NED meningkat serta CER dan WER dapat ditekan.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berhasil mengembangkan sistem OCR untuk pembacaan daftar bahan pada label makanan kemasan melalui integrasi model CRAFT pada tahap deteksi dan model CRNN berarsitektur TRBA pada tahap pengenalan. Sistem dirancang melalui rangkaian proses yang meliputi deteksi area teks, pemotongan sub-citra kata, pengenalan karakter, dan rekonstruksi urutan baca. Hasil implementasi menunjukkan bahwa pendekatan *end-to-end* berbasis dua tahap dapat diterapkan secara operasional pada domain teks label makanan.

Hasil pengujian pada data uji menunjukkan bahwa proses *fine-tuning* memberikan peningkatan performa yang konsisten dibandingkan model pra-latih. Nilai *Word Accuracy* meningkat dari 90,381% menjadi 94,693%, NED meningkat dari 0,946 menjadi 0,987, CER menurun dari 0,066 menjadi 0,015, dan WER menurun dari 0,096 menjadi 0,053. Perubahan metrik tersebut menegaskan bahwa model yang telah disesuaikan lebih mampu mengikuti karakteristik teks daftar bahan pada kemasan, baik pada tingkat kata maupun karakter. Selain itu, pada tahap implementasi sistem juga mampu menghasilkan keluaran teks daftar bahan yang lebih terstruktur dibanding keluaran kata terpisah, sehingga proses pembacaan komposisi menjadi lebih mudah ditinjau dan dimanfaatkan untuk kebutuhan lanjutan berbasis teks. Temuan ini memperlihatkan bahwa kombinasi strategi deteksi yang tepat, adaptasi model pengenalan, dan rekonstruksi urutan baca merupakan komponen penting dalam peningkatan kualitas OCR pada domain label makanan.

Secara keseluruhan, penelitian ini diharapkan dapat memberikan fondasi yang lebih kuat bagi pengembangan sistem otomatis pembacaan komposisi bahan pada label makanan kemasan, baik untuk kebutuhan akademik maupun implementasi praktis. Secara akademik, hasil penelitian ini memperkuat bukti bahwa adaptasi model OCR pada domain data spesifik dapat meningkatkan kinerja pengenalan teks. Secara praktis, sistem yang dikembangkan berpotensi menjadi dasar bagi layanan lanjutan yang membutuhkan ekstraksi teks komposisi secara cepat, konsisten, dan terstruktur.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, penulis memiliki beberapa saran untuk pengembangan penelitian selanjutnya:

- a. Memperluas jumlah dan variasi data label makanan kemasan agar model lebih kuat terhadap perbedaan kualitas citra, tata letak teks, dan karakteristik kemasan.
- b. Mengembangkan sistem menjadi aplikasi skrining komposisi bahan untuk mendukung pengecekan status halal berbasis kamus atau aturan bahan.
- c. Menambahkan modul identifikasi alergen otomatis dari hasil OCR seperti susu, kedelai, kacang, gluten.
- d. Memperluas dukungan multibahasa pada label kemasan agar sistem dapat dipakai lintas produk impor.
- e. Menambahkan fitur normalisasi dan koreksi teks pasca-OCR untuk mengurangi kesalahan karakter minor.
- f. Mengintegrasikan sistem dengan basis data bahan pangan agar interpretasi hasil lebih informatif.
- g. Mengembangkan antarmuka menjadi aplikasi *mobile* agar proses pemindaian label bisa dilakukan langsung di lapangan.

DAFTAR PUSTAKA

- Baek, J., Kim, G., Lee, J., Park, S., Han, D., Yun, S., Oh, S. J., & Lee, H. (2019). What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis. *Proceedings of the IEEE International Conference on Computer Vision*, 4714–4722. <http://arxiv.org/abs/1904.01906>
- Baek, Y., Lee, B., Han, D., Yun, S., & Lee, H. (2019). *Character Region Awareness for Text Detection*. <https://doi.org/10.1109/CVPR.2019.00959>
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). *NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE*.
- Buga, R., Buzea, C. G., Agop, M., Ochiuz, L., Vasincu, D., Popa, O., Rusu, D. I., Știrban, I., & Eva, L. (2025). Streamlit Application and Deep Learning Model for Brain Metastasis Monitoring After Gamma Knife Treatment. *Biomedicines* 2025, Vol. 13, Page 423, 13(2), 423. <https://doi.org/10.3390/BIOMEDICINES13020423>
- Chu, H. (2011). *MDB: A Memory-Mapped Database and Backend for OpenLDAP*. <http://www.symas.com>, <http://www.openldap.org>
- Guimarães, V., Nascimento, J., Viana, P., & Carvalho, P. (2023). A Review of Recent Advances and Challenges in Grocery Label Detection and Recognition. *Applied Sciences* 2023, Vol. 13, Page 2871, 13(5), 2871. <https://doi.org/10.3390/APP13052871>
- Gupta, A., Vedaldi, A., & Zisserman, A. (2016). Synthetic Data for Text Localisation in Natural Images. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December*, 2315–2324. <https://doi.org/10.1109/CVPR.2016.254>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Ingredients Image from Food label*. (n.d.). Retrieved April 6, 2026, from <https://www.kaggle.com/datasets/shensivam/ingredients-image-from-food-label>
- Jaderberg, M., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). *Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition*. <https://arxiv.org/pdf/1406.2227>
- Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., & Wei, F. (2021). TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models.

- Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023*, 37, 13094–13102. <https://doi.org/10.1609/aaai.v37i11.26538>
- Magadán, L., Granda, J. C., & Suárez, F. J. (2024). Robust prediction of remaining useful lifetime of bearings using deep learning. *Engineering Applications of Artificial Intelligence*, 130. <https://doi.org/10.1016/J.ENGAPPAL.2023.107690>
- Manuel, T., Saputra, E., Susanto, A., & Carmelita, J. (2024). Implementation of Tesseract OCR and Bounding Box for Text Extraction on Food Nutrition Labels. *Building of Informatics, Technology and Science (BITS)*, 6(3), 1403–1412. <https://doi.org/10.47065/BITS.V6I3.6107>
- Nagayi, M., Khan, A., Frank, T., Swart, R., & Nyirenda, C. (2025). *Evaluating OCR performance on food packaging labels in South Africa*. <https://doi.org/10.48550/arXiv.2510.03570>
- Neudecker, C., Baier, K., Gerber, M., Christian, C., Apostolos, A., & Stefan, P. (2021). A survey of OCR evaluation tools and metrics. *ACM International Conference Proceeding Series*, 13–18. <https://doi.org/10.1145/3476887.3476888>
- Pan, S. J., & Yang, Q. (2009). *A Survey on Transfer Learning*. <https://doi.org/10.1109/TKDE.2009.191>
- Peng, H., Bayon, J., Recas, J., & Guijarro, M. (2025). Efficient Expiration Date Recognition in Food Packages for Mobile Applications. *Algorithms*, 18(5). <https://doi.org/10.3390/A18050286>
- Rosyadi, A. W., Ma'shumah, S., Zaman, M. Q., & Fajar, M. R. (2024). Ingredients Identification Through Label Scanning Using PaddleOCR and ChatGPT for Information Retrieval. *Jurnal RESTI*, 8(6), 758–767. <https://doi.org/10.29207/RESTI.V8I6.6119>
- Seitaj, H., & Elangovan, V. (2024). *Information Extraction from Product Labels: A Machine Vision Approach*. 227–247. <https://doi.org/10.5121/csit.2024.141419>
- Shi, B., Bai, X., & Yao, C. (2015). An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11), 2298–2304. <https://doi.org/10.1109/TPAMI.2016.2646371>
- Shi, B., Wang, X., Lyu, P., Yao, C., & Bai, X. (2016). Robust Scene Text Recognition with Automatic Rectification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, 4168–4176. <https://doi.org/10.1109/CVPR.2016.452>

- Wang, X.-F., He, Z.-H., Wang, K., Wang, Y.-F., Zou, L., & Wu, Z.-Z. (2023). A survey of text detection and recognition algorithms based on deep learning technology. *Neurocomputing*, 556, 126702. <https://doi.org/https://doi.org/10.1016/j.neucom.2023.126702>
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). *How transferable are features in deep neural networks?*

LAMPIRAN