

**IMPLEMENTASI *OPTICAL CHARACTER
RECOGNITION* DENGAN ANALISIS TATA
LETAK UNTUK REKONTRUKSI DOKUMEN**

Proposal Tugas Akhir

Oleh

**Habib Akhmad Al Farisi
18222029**



**PROGRAM STUDI SISTEM DAN TEKNOLOGI INFORMASI
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
Nopember 2025**

LEMBAR PENGESAHAN

IMPLEMENTASI *OPTICAL CHARACTER RECOGNITION* DENGAN ANALISIS TATA LETAK UNTUK REKONSTRUKSI DOKUMEN

Proposal Tugas Akhir

Oleh

**Habib Akhmad Al Farisi
18222029**

Program Studi Sistem dan Teknologi Informasi
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung

Proposal Tugas Akhir ini telah disetujui dan disahkan
di Bandung, pada tanggal 4 Nopember 2025

Pembimbing

Dr. Riza Satria Perdana, S.T, M.T.
NIP. 197006091995121002

DAFTAR ISI

| | |
|--|------------|
| DAFTAR GAMBAR | iii |
| DAFTAR TABEL | iv |
| DAFTAR KODE | v |
| I PENDAHULUAN | 1 |
| I.1 Latar Belakang | 1 |
| I.2 Rumusan Masalah | 2 |
| I.3 Tujuan | 2 |
| I.4 Batasan Masalah | 2 |
| I.5 Metodologi | 3 |
| II STUDI LITERATUR | 5 |
| II.1 Digitalisasi Dokumen | 5 |
| II.2 <i>Optical Character Recognition</i> (OCR) | 6 |
| II.2.1 SEJARAH PERKEMBANGAN OCR | 6 |
| II.2.2 Jenis-Jenis Sistem OCR | 9 |
| II.2.3 Tahapan Utama dalam Proses OCR | 10 |
| II.3 <i>Document Layout Analysis</i> (DLA) | 11 |
| III ANALISIS MASALAH | 16 |
| III.1 Analisis Kondisi Saat Ini | 16 |
| III.2 Analisis Kebutuhan | 16 |
| III.2.1 Identifikasi Masalah Pengguna | 17 |
| III.2.2 Kebutuhan Fungsional | 17 |
| III.2.3 Kebutuhan Nonfungsional | 18 |
| III.3 Analisis Pemilihan Solusi | 19 |
| III.3.1 Alternatif Solusi | 19 |
| III.3.1.1 Kombinasi YOLO, CRNN, dan <i>Post-processing</i> | 20 |
| III.3.1.2 PaddleOCR | 20 |
| III.3.1.3 Kombinasi EasyOCR dan Layout Parser | 21 |
| III.3.2 Analisis Penentuan Solusi | 22 |
| IV DESAIN KONSEP SOLUSI | 27 |
| V RENCANA SELANJUTNYA | 28 |

DAFTAR GAMBAR

| | | |
|------|---|----|
| II.1 | Tipe-Tipe OCR (Islam, Islam, dan Noor 2017) | 9 |
| II.2 | Contoh Hasil DLA | 12 |
| II.3 | Kerangka Kerja DLA secara Umum (Binmakhashen dan Mahmoud 2019) | 13 |

DAFTAR TABEL

| | | |
|-------|--|----|
| III.1 | Daftar Kebutuhan Fungsional Sistem | 18 |
| III.2 | Daftar Kebutuhan Nonfungsional Sistem | 19 |
| III.3 | Kelebihan dan Kekurangan Masing-Masing Alternatif Solusi | 23 |
| III.5 | Daftar Kebutuhan Nonfungsional Sistem | 25 |

DAFTAR KODE

BAB I

PENDAHULUAN

I.1 Latar Belakang

Menurut Rihanna, Eryanto, dan Monoarfa (2024), keberadaan arsip digital mempermudah peneliti dan masyarakat dalam mengakses informasi sesuai dengan kebutuhan mereka dari mana saja. Kondisi ini menunjukkan pentingnya proses pengubahan dokumen fisik menjadi dokumen digital yang dapat disunting dan dicari, terutama untuk meningkatkan efisiensi pengelolaan arsip serta memudahkan distribusi informasi. Namun, proses konversi tersebut masih menghadapi berbagai tantangan, terutama pada dokumen yang memiliki struktur kompleks, seperti dokumen dengan elemen tabel, gambar, atau tulisan tangan. Teknologi *Optical Character Recognition* (OCR) memang mampu mengenali teks pada dokumen fisik, tetapi sering kali gagal mempertahankan tata letak aslinya (Tripathi 2025a), sehingga dokumen digital yang dihasilkan menjadi tidak terstruktur dengan baik dan berbeda dari tampilan dokumen fisiknya.

Untuk mengatasi keterbatasan tersebut, diperlukan pendekatan OCR yang tidak hanya mampu mengenali karakter, tetapi juga memahami tata letak dokumen secara keseluruhan. Menurut Wang dkk. (2023), *Document Layout Analysis* (DLA) atau analisis tata letak dokumen merupakan proses mendeteksi berbagai komponen semantik dalam suatu dokumen dan mengklasifikasikannya ke dalam kategori seperti teks, judul, tabel, atau gambar. Dengan menerapkan DLA, sistem dapat memahami hubungan antar elemen dokumen, sehingga hasil digitalisasi menjadi lebih akurat dan tetap menyerupai format aslinya.

Berbagai penelitian telah berupaya meningkatkan akurasi DLA melalui pendekatan *deep learning*. Model berbasis *Convolutional Neural Network* (CNN) seperti Faster R-CNN dan Mask R-CNN menunjukkan kinerja yang baik dalam mendeteksi elemen tata letak, terutama saat dilatih menggunakan dataset berskala besar seperti

PubLayNet (Zhong, Tang, dan Jimeno Yepes 2019). Seiring dengan kemajuan teknologi, pendekatan berbasis *Transformer* mulai digunakan untuk memahami konteks dokumen secara lebih menyeluruh. Model seperti LayoutLM (Xu dkk. 2020) dan DiT (Li dkk. 2022) mampu menggabungkan representasi visual dan tekstual secara terpadu, sehingga meningkatkan kemampuan sistem dalam menganalisis dokumen yang kompleks.

Berdasarkan kondisi tersebut, penelitian ini berfokus pada pengembangan sistem yang mengintegrasikan kemampuan OCR dengan analisis tata letak dokumen berbasis *deep learning*. Sistem yang diusulkan bertujuan menghasilkan dokumen digital yang rapi, terstruktur, dan dapat disunting dari sumber dokumen fisik. Penelitian ini diharapkan dapat memberikan kontribusi terhadap peningkatan akurasi proses konversi dokumen, terutama dalam mempertahankan tata letak aslinya.

I.2 Rumusan Masalah

Berdasarkan identifikasi yang telah dilakukan, berikut adalah rumusan masalah yang akan menjadi fokus dalam tugas akhir ini.

1. Bagaimana menerapkan OCR untuk mengekstraksi teks, tabel, dan gambar dari dokumen fisik?
2. Bagaimana merancang sistem OCR yang mampu mempertahankan tata letak asli dokumen fisik sehingga dokumen digital hasil OCR memiliki struktur yang rapi dan dapat disunting?

I.3 Tujuan

Berdasarkan rumusan masalah yang ada, tujuan dari tugas akhir ini adalah mengembangkan sistem OCR yang mampu menghasilkan dokumen digital dari dokumen tercetak maupun tulisan tangan dengan tetap mempertahankan tata letak asli dokumennya. Sistem ini tidak hanya mengenali teks, tetapi juga mampu mendeteksi elemen seperti tabel dan gambar serta menempatkannya kembali dalam posisi yang sesuai dengan tata letak pada dokumen sumber, sehingga menghasilkan dokumen digital yang terstruktur, rapi, dan dapat disunting.

I.4 Batasan Masalah

Berikut disajikan batasan masalah yang digunakan untuk memperjelas ruang lingkup pembahasan dalam tugas akhir ini.

1. Jenis dokumen yang diproses terbatas pada dokumen fisik berbahasa Indonesia.
2. Elemen dokumen yang dikenali dan direkonstruksi meliputi teks, tabel, dan gambar.
3. Sistem dirancang untuk memproses dokumen yang memiliki struktur tata letak satu kolom, yaitu dokumen dengan alur baca berurutan dari atas ke bawah tanpa pembagian kolom.
4. Pengaturan *font* pada hasil dokumen digital tidak meniru *font* asli dokumen sumber.
5. Sistem tidak dirancang untuk mengenali atau memproses dokumen yang mengalami kerusakan berat, memiliki orientasi terputar, dan tercoret.
6. Keluaran sistem berupa dokumen digital yang memiliki latar belakang putih polos, tanpa membawa latar belakang asli dari dokumen sumber.

I.5 Metodologi

Metodologi penelitian dalam tugas akhir ini mencakup lima tahapan utama, yakni eksplorasi, perancangan, implementasi, pengujian, dan evaluasi. Alur penelitian ini pada dasarnya berjalan secara berurutan. Namun, untuk memastikan hasil yang optimal, diterapkan mekanisme iteratif. Apabila hasil pada tahap pengujian belum memuaskan, proses dapat diulang kembali. Iterasi ini umumnya dilakukan mulai dari tahap perancangan dan implementasi, namun tidak menutup kemungkinan pengulangan kembali ke tahap eksplorasi jika metode atau teknologi yang dipilih terbukti tidak memadai.

1. Eksplorasi OCR dan metode analisis tata letak dokumen

Tahap pertama yang dilakukan adalah eksplorasi, yang berfokus pada pemilihan teknologi OCR dan metode analisis tata letak dokumen. Pada tahap ini dilakukan studi mendalam terhadap berbagai teknologi OCR yang tersedia, dengan tujuan untuk menentukan solusi terbaik dalam mengubah dokumen fisik menjadi dokumen digital secara akurat. Selain menilai kemampuan OCR dalam mengenali teks, tahap eksplorasi juga mencakup analisis terhadap cara setiap OCR menangani berbagai elemen dokumen, seperti tabel dan gambar. Di samping itu, dipelajari pula metode yang dapat digunakan untuk mendeteksi dan memetakan posisi elemen-elemen tersebut sehingga dokumen digital yang dihasilkan tidak mampu mempertahankan tata letak aslinya.

2. Perancangan

Setelah teknologi OCR dan metode analisis tata letak dokumen ditentukan, dilakukan tahap perancangan arsitektur sistem. Pada tahap ini, dirancang me-

kanisme utama yang berfungsi untuk mengintegrasikan hasil ekstraksi dari OCR dengan informasi posisi setiap elemen pada dokumen fisik. Tujuannya adalah agar sistem mampu merekonstruksi dokumen digital yang tidak hanya menyalin isi dokumen sumber, tetapi juga mempertahankan tata letak aslinya.

3. Implementasi

Setelah tahap perancangan selesai dilakukan, proses dilanjutkan ke tahap implementasi yang merupakan bagian inti dari penelitian ini. Pada tahap ini, rancangan mekanisme rekonstruksi dokumen yang telah disusun sebelumnya mulai diterapkan dalam bentuk pengembangan sistem secara nyata. Implementasi dilakukan dengan menulis kode program yang berfungsi untuk memproses dokumen fisik, menerapkan teknologi OCR untuk mengekstraksi elemen dokumen, serta mengintegrasikannya dengan hasil analisis tata letak agar struktur halaman tetap terjaga.

4. Pengujian

Setelah tahap implementasi selesai dilakukan, proses penelitian dilanjutkan dengan tahap pengujian. Pada tahap ini, dilakukan serangkaian uji coba untuk menilai kinerja sistem dalam mempertahankan tata letak dokumen. Pengujian dilakukan dengan membandingkan hasil yang hanya menggunakan OCR dengan hasil dokumen yang telah melalui mekanisme penyusunan ulang tata letak. Melalui perbandingan ini, dapat dinilai kemampuan sistem untuk menghasilkan dokumen digital yang rapi dan menyerupai bentuk aslinya. Apabila hasil yang diperoleh belum memuaskan, maka tahap eksplorasi, perancangan, implementasi, dan pengujian dapat diulang secara iteratif hingga sistem memberikan hasil yang optimal.

5. Evaluasi

Tahap terakhir dalam penelitian ini adalah evaluasi. Pada tahap ini dilakukan analisis terhadap hasil pengujian untuk menilai efektivitas sistem dalam mencapai tujuan penelitian. Evaluasi mencakup peninjauan terhadap keterbacaan hasil dan kesesuaian dengan tata letak dokumen fisik. Berdasarkan hasil evaluasi tersebut, disusun kesimpulan yang bersifat ringkas dan jelas guna menjawab rumusan masalah serta memberikan gambaran mengenai kemampuan sistem OCR yang dikembangkan dalam mempertahankan tata letak dokumen dengan baik.

BAB II

STUDI LITERATUR

II.1 Digitalisasi Dokumen

Digitalisasi dokumen didefinisikan sebagai proses pengubahan data dari format analog (seperti dokumen fisik) menjadi format digital yang dapat dibaca dan diproses oleh komputer. Proses ini sangat penting dalam transformasi digital karena melibatkan pengkodean potongan data menjadi format digital, yang memungkinkan transmisi, penggunaan kembali, dan pemrosesan informasi secara efisien (Tan, Chi, dan Lam 2023). Dalam konteks dokumen, digitalisasi berfokus pada pengelolaan sejumlah besar informasi yang umumnya tidak terstruktur, seperti hasil pindai, PDF, dan berkas digital asli yang belum dapat dibaca atau diproses langsung oleh komputer (Sinha dan S 2025).

Salah satu teknologi kunci dalam proses digitalisasi dokumen adalah *Optical Character Recognition* (OCR). OCR adalah ilmu yang memungkinkan terjemahan berbagai jenis dokumen atau gambar menjadi data yang dapat dianalisis, diedit, dan dicari (Memon dkk. 2020). Dalam beberapa dekade terakhir, peneliti telah memanfaatkan alat kecerdasan buatan (*artificial intelligence*) dan pembelajaran mesin (*machine learning*) untuk menganalisis dokumen tulisan tangan dan cetak secara otomatis, mengubahnya menjadi format elektronik yang dapat digunakan (Memon dkk. 2020).

Digitalisasi dokumen, yang didukung oleh implementasi teknologi modern, menawarkan berbagai manfaat penting. Manfaat-manfaat tersebut akan diuraikan sebagai berikut:

1. Aksesibilitas dan efisiensi pencarian

Dokumen yang telah didigitasi dan diproses dengan teknologi seperti OCR diubah menjadi data yang dapat dicari. Hal ini secara signifikan memudahkan pengambilan informasi yang dibutuhkan, menghilangkan kebutuhan untuk

mencari di tumpukan dokumen fisik (Sinha dan S 2025).

2. **Preservasi dan perlindungan**

Digitalisasi memainkan peran penting dalam pelestarian data, terutama untuk data historis dan arsip. Dengan mengalihmediakan dokumen fisik menjadi format digital, hal ini membantu melestarikan informasi dari kerusakan akibat seringnya penggunaan, atau faktor lingkungan (seperti kelembaban dan suhu) maupun faktor biologis (seperti hama perusak) (Rihanna, Eryanto, dan Monoarfa 2024)

3. **Skalabilitas dan otomatisasi proses data**

Pengumpulan data secara tradisional, misalnya melalui transkripsi manual, sangat terbatas. Metode ini tidak efisien untuk menangani data berskala besar, memakan biaya, dan prosesnya rumit. Oleh karena itu, *Machine Learning* (ML) modern kini digunakan untuk mengotomatisasi digitalisasi. Solusi ML terbukti lebih cepat, mudah diskalakan untuk jutaan dokumen, dan hasilnya dapat diandalkan secara konsisten. (Dahl dkk. 2021)

4. **Peningkatan kualitas analisis data dan penelitian**

Ketersediaan data hasil digitalisasi sangat menentukan mutu penelitian. Dengan kumpulan data yang lebih besar dan rinci, peneliti dapat menarik kesimpulan yang lebih tajam. Hal ini memungkinkan mereka menjawab pertanyaan penelitian yang kompleks secara lebih meyakinkan (Dahl dkk. 2021). Selain itu, teks yang diekstrak dari dokumen digital dapat dianalisis oleh *Large Language Model* (LLM) untuk mengidentifikasi pasangan kunci-nilai dan mengatasi ambiguitas, sehingga mampu menyajikan data secara terstruktur berdasarkan konteks informasinya (Sinha dan S 2025).

II.2 *Optical Character Recognition* (OCR)

Menurut Islam, Islam, dan Noor (2017), OCR adalah perangkat lunak yang mengubah teks dan gambar tercetak menjadi bentuk digital sehingga dapat digunakan oleh mesin. Borovikov (2014) menyatakan bahwa OCR adalah proses mengubah gambar yang dipindai dari teks tercetak atau tulisan tangan (angka, huruf, dan simbol) menjadi teks yang dapat dibaca mesin, baik dalam bentuk berkas teks biasa maupun format HTML.

II.2.1 SEJARAH PERKEMBANGAN OCR

Teknologi OCR telah berevolusi secara signifikan. Perjalanannya dimulai dari sistem mekanis sederhana di era awal, yang kini telah bertransformasi menjadi sis-

tem canggih berbasis *deep learning*. Berikut adalah linimasa sejarah perkembangan OCR menurut Tripathi (2025b).

1. Konsep awal (1920-an)

Era 1920-an menjadi fondasi awal OCR, dimulai oleh Emanuel Goldberg yang menciptakan sistem temu balik dokumen elektronik pertama di dunia, dikenal sebagai Mesin Statistik. Sistem ini menggunakan sel fotolistrik dan proyektor film untuk mengatasi kesulitan menemukan catatan keuangan pada mikrofilm, serta mampu menyortir surat dan membaca cek bank berdasarkan pola visual. Prinsip detektor fotoelektrik Goldberg inilah yang kemudian diadaptasi pada tahun 1929 oleh Gustav Tauschek, seorang penemu otodidak asal Austria, yang mengembangkan Mesin Pembaca Analog. Mesin Tauschek bekerja dengan cara memindai gambar berteks melalui sebuah jendela, kemudian menggunakan cakram berputar berisi potongan huruf dan angka untuk mencocokkan karakter. Setelah karakter dikenali, mesin akan secara otomatis mengaktifkan drum pencetak untuk mencetak teks hasil pengenalan karakter tersebut.

2. Mesin OCR pertama (1950-an)

Era 1950-an, yang didorong oleh pertumbuhan data sangat pesat, menandai lahirnya kebutuhan akan teknologi pemrosesan data yang lebih efisien. Menjawab tantangan ini, David Shepard dan Harvey Cook Jr. menciptakan alat bernama GISMO, yang kemudian dipatenkan dengan nama Analyzing Reader. Alat ini merevolusi teknik pengambilan data otomatis berkat kemampuannya mengubah teks tercetak menjadi bahasa atau kode mesin. Untuk mengkomersialkan penemuan tersebut, Shepard bersama Cook dan William Lawless Jr. mendirikan Intelligent Machines Research Co. (IMR) pada tahun 1952. Mesin hasil pengembangan mereka segera diadopsi oleh berbagai perusahaan besar, seperti Reader's Digest, National City Bank, dan beberapa perusahaan minyak utama. Puncaknya terjadi pada tahun 1959 setelah memperoleh lisensi paten dari IMR, IBM memperkenalkan sistem baru untuk menangkap data dari dokumen. IBM menamai sistem inilah yang kini dikenal sebagai *Optical Character Recognition* (OCR), dan istilah tersebut kemudian dengan cepat menjadi standar industri.

3. Kemajuan pengenalan pola (1960-an hingga 1970-an)

Pada 1960-an, perkembangan OCR mengalami kemajuan pesat. Penelitian di MIT berfokus pada peningkatan perangkat lunak serta penyederhanaan proses pengambilan dan penyimpanan data. Peneliti di MIT mengembangkan algoritma yang mampu menganalisis data dan menyesuaikan diri dengan ber-

bagai format dokumen. Pengembangan inilah yang kemudian menjadi dasar bagi lahirnya *machine learning* dalam OCR. Pada periode yang sama, muncul Teknik Transformasi Hough (*Hough Transform*) yang memungkinkan mesin OCR mengenali tidak hanya huruf, tetapi juga bentuk geometris. Dua teknologi penting kemudian lahir sebagai spesialisasi yaitu *Intelligent Character Recognition* (ICR) untuk mengenali tulisan tangan, dan *Magnetic Ink Character Recognition* (MICR) untuk memproses cek secara otomatis di industri perbankan. Kehadiran ICR dan MICR ini menandai kemampuan OCR untuk beradaptasi dan memperluas penerapannya di berbagai bidang.

4. Kemajuan OCR digital (1980-an hingga 1990-an)

Era 1980–1990 menandai pergeseran fundamental dalam OCR, fokus pengembangan beralih secara masif dari perangkat keras ke perangkat lunak. Pergeseran ini dipicu oleh adopsi teknologi pencitraan digital yang membuka potensi besar untuk otomatisasi data dalam skala yang lebih luas. Berkat fokus pada perangkat lunak, lahirlah algoritma yang jauh lebih canggih. Sistem OCR kini mampu mengenali berbagai jenis *font* dan memahami tata letak dokumen yang kompleks. Kemampuan ini secara langsung mempercepat proses entri data otomatis sekaligus meningkatkan akurasi pengenalan tulisan tangan. Selain itu, *Document Layout Analysis* (DLA) memungkinkan ekstraksi data yang lebih akurat dari elemen seperti tabel dan grafik, serta dukungan multibahasa yang mulai berkembang untuk memenuhi kebutuhan perusahaan global, menjadikan OCR solusi yang jauh lebih praktis dan dapat diterapkan secara luas.

5. Perangkat lunak OCR komersial (1990-an)

Era 1990-an menjadi titik komersialisasi OCR secara massal. Perangkat lunak dari perusahaan seperti ABBYY, Adobe, dan Nuance mulai mendominasi pasar, memudahkan pengguna mengonversi dokumen cetak menjadi teks yang dapat diedit. Kemampuan ini secara langsung meningkatkan efisiensi bisnis dan mendorong era digitalisasi arsip berskala besar.

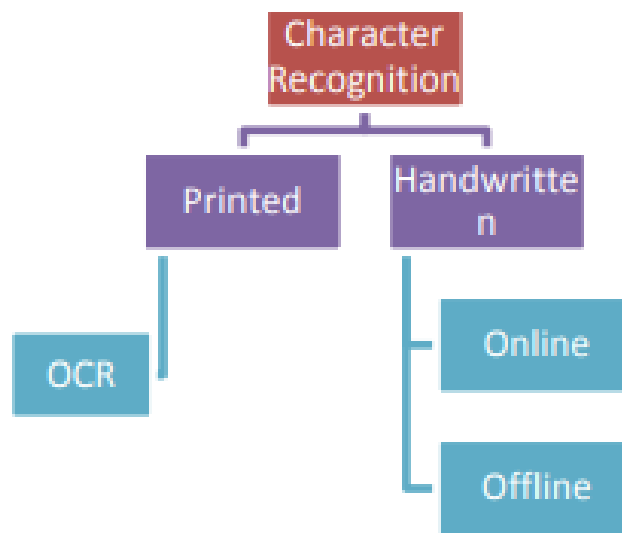
6. Tesseract OCR (2005)

Tahun 2005 menjadi titik penting ketika Tesseract OCR dihidupkan kembali oleh Google sebagai proyek *open-source*. Dengan memanfaatkan *machine learning* (ML) dan *computer vision* (CV), Tesseract mampu meningkatkan akurasi ekstraksi teksnya secara signifikan, hingga mencapai lebih dari 70%. Selain itu, statusnya sebagai proyek *open-source* untuk pertama kalinya memungkinkan kolaborasi dan kontribusi dari komunitas pengembang global.

7. Revolusi *deep learning* (2010-an hingga Sekarang)

Memasuki 2010-an, revolusi *deep learning* meningkatkan akurasi OCR hingga mendekati 99%. Kemajuan ini didorong oleh dua arsitektur utama: *Convolutional Neural Networks* (CNN) yang unggul dalam ekstraksi fitur gambar dan pengenalan *font* kompleks, serta *Recurrent Neural Networks* (RNN) yang berperan besar dalam meningkatkan pemahaman konteks dan akurasi pengenalan tulisan tangan. Kombinasi kedua teknologi inilah yang mengubah OCR secara fundamental, yang hasilnya membuat interpretasi teks tulisan tangan menjadi jauh lebih akurat dan sistem mampu menangani struktur dokumen yang kompleks dengan lebih baik.

II.2.2 Jenis-Jenis Sistem OCR



Gambar II.1 Tipe-Tipe OCR (Islam, Islam, dan Noor 2017)

Pengkategorian paling mendasar dari sistem OCR didasarkan pada jenis input yang diproses (Islam, Islam, dan Noor 2017). Berdasarkan hal ini, OCR dibagi menjadi dua jenis utama.

1. *Printed text recognition*

Sistem yang dirancang untuk mengenali karakter yang dicetak oleh mesin, yang umumnya memiliki dimensi seragam.

2. *Handwriting recognition*

Sistem yang mengenali tulisan tangan manusia. Jenis ini dianggap sebagai tugas yang jauh lebih sulit karena tingginya perbedaan gaya penulisan antar individu, atau bahkan oleh individu yang sama.

Lebih lanjut, Islam, Islam, dan Noor (2017) membagi pengenalan tulisan tangan

menjadi dua subkategori berdasarkan mode perolehan data.

1. *Online recognition*

Dilakukan secara *real-time* saat pengguna menulis. Sistem ini dapat menangkap informasi temporal seperti kecepatan, arah goresan, dan jumlah goresan, sehingga membuatnya lebih mudah dikembangkan.

2. *Offline recognition*

Bekerja pada data statis, yaitu gambar *bitmap* dari tulisan yang sudah ada. Ini adalah masalah yang jauh lebih sulit karena tidak ada informasi temporal yang tersedia untuk membantu proses pengenalan.

Sementara itu, Borovikov (2014) mengklasifikasikan sistem OCR komersial berdasarkan ruang lingkup dan tujuannya.

1. *Task specific readers*

Sistem ini dirancang untuk menangani tipe dokumen yang sangat spesifik dengan layout yang sudah ditentukan, seperti formulir standar, cek bank, atau slip kartu kredit. Fokus utama sistem ini adalah kecepatan pemrosesan yang sangat tinggi dan tingkat kesalahan yang sangat rendah.

2. *General purpose page readers*

Sistem ini dirancang untuk menangani rentang dokumen yang lebih luas dan beragam, seperti surat bisnis, karya ilmiah, dan surat kabar.

II.2.3 Tahapan Utama dalam Proses OCR

Proses OCR dilakukan melalui beberapa fase utama. Berikut merupakan fase-fase utama pada OCR menurut Islam, Islam, dan Noor (2017).

1. *Image acquisition*

Tahap awal dalam OCR adalah memperoleh gambar digital dari sumber eksternal (seperti scanner atau kamera). Gambar ini kemudian diubah ke format yang dapat diproses komputer, melibatkan kuantisasi dan binarisasi untuk menyederhanakan intensitas piksel. Gambar juga dapat dikompresi untuk mengurangi ukuran berkas.

2. *Preprocessing*

Langkah ini bertujuan meningkatkan kualitas visual gambar untuk pemrosesan lebih lanjut. Kegiatan utamanya meliputi penghilangan *noise*, memisahkan teks dan latar belakang (*thresholding*), koreksi kemiringan (*deskew*), dan operasi morfologi (erosi, dilasi). Proses *thinning* dan ekstraksi garis dasar (*baseline*) juga diterapkan untuk mempermudah pemrosesan karakter.

3. *Character segmentation*

Fase ini berfungsi untuk memisahkan teks menjadi satuan karakter indivi-

dual agar dapat dikenali. Proses segmentasi dapat dilakukan secara eksplisit maupun implisit dengan menggunakan teknik, seperti analisis komponen terhubung atau profil proyeksi. Ketepatan segmentasi sangat penting karena berpengaruh langsung terhadap keberhasilan tahap klasifikasi berikutnya.

4. *Feature extraction*

Setelah karakter dipisahkan, tahap ini bertujuan mengekstraksi fitur-fitur unik yang membedakan satu karakter dari karakter lainnya. Fitur dapat bersifat geometris (seperti *loop* atau *stroke*) maupun statistik (seperti *moments*). Pemilihan fitur yang tepat sangat penting untuk meminimalkan variasi antarkarakter sejenis. Untuk meningkatkan efisiensi, teknik seperti *Principal Component Analysis* (PCA) dapat digunakan.

5. *Character classification*

Fase ini merupakan tahap pengenalan, yaitu proses ketika fitur karakter ditetapkan ke dalam kelas atau kategori tertentu (misalnya huruf 'A', 'B', dan seterusnya). Metode klasifikasi dapat bersifat struktural, yang didasarkan pada hubungan antar komponen, atau bersifat statistik, yang menggunakan pendekatan seperti Bayesian dan jaringan saraf (*neural network*). Keberhasilan proses klasifikasi sangat bergantung pada kualitas segmentasi serta fitur yang diekstraksi pada tahap-tahap sebelumnya.

6. Post Processing

Tahap terakhir ini bertujuan untuk meningkatkan akurasi hasil OCR setelah proses klasifikasi. Perbaikan dilakukan dengan menggunakan kombinasi *classifier* secara paralel maupun berurutan. Selain itu, diterapkan analisis konteks secara geometris, linguistik, maupun dokumen, serta pemanfaatan kamus, *spell checker*, dan model probabilistik (seperti *Markov* atau *n-grams*) untuk memperbaiki kesalahan transkripsi, sehingga menghasilkan keluaran yang lebih akurat dan konsisten.

II.3 *Document Layout Analysis (DLA)*

Document Layout Analysis (DLA) merupakan langkah awal yang krusial dalam sistem pemahaman dokumen (Binmakhashen dan Mahmoud 2019). Tugas utama DLA adalah mendeteksi dan melakukan anotasi terhadap struktur fisik dokumen (Binmakhashen dan Mahmoud 2019). Secara lebih spesifik, DLA melibatkan pemahaman terhadap pengaturan elemen-elemen di dalam suatu dokumen (Shehzadi, Stricker, dan Afzal 2024).

Tujuan utama dari DLA adalah mempermudah proses analisis atau pengenalan pada

tahap berikutnya dengan cara mengidentifikasi blok-blok dokumen yang memiliki keseragaman serta menentukan hubungan antarblok tersebut (Binmakhshen dan Mahmoud 2019). Dengan kata lain, DLA berperan sebagai komponen kunci dalam *document intelligence* karena kemampuannya mengubah dokumen yang tidak terstruktur menjadi format yang terstruktur. Hal ini sangat penting untuk mendukung proses identifikasi dan ekstraksi data dari dokumen (Shehzadi, Stricker, dan Afzal 2024).

abandon:0.906
SYMMETRY

isolate_formula:0.696
plain text:0.925 $OC \times OC' = k^2 = OA \times OA'$
and
isolate_formula:0.931
plain text:0.942 $r^2 = AO \times AC' = OA \times C'A$
(in the notation of directed distances), we have
isolate_formula:0.967

$$\frac{OC}{OA'} = \frac{OA}{OC'} = \frac{OA}{OA - C'A} = \frac{OA^2}{OA^2 - (OA \times C'A)}$$

$$= \frac{OA^2}{OA^2 - r^2} = \frac{\epsilon^2}{\epsilon^2 - 1},$$
plain text:0.979
which is negative or positive according as $\epsilon < 1$ or $\epsilon > 1$. Hence, for an ellipse the center C and directrix a are on opposite sides of O , as in Figure 6.6B, but for a hyperbola they are on the same side, as in Figure 6.6C. In other words, the ellipse encloses its two foci and lies entirely between its two directrices, but the two directrices of a hyperbola both lie in the "empty" space between the two branches.

figure:0.971

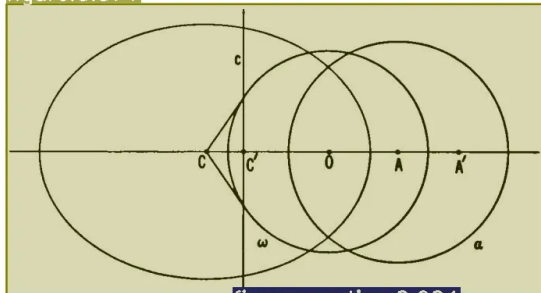


figure-caption:0.904
Figure 6.6D

plain text:0.980
In mechanics we learn that, when air resistance is neglected, the trajectory of a thrown ball is an arc of a parabola whose focus can be located without much difficulty. Since the thrown ball is, for a few seconds, a little artificial satellite, the apparent parabola is more accurately an enormously elongated ellipse, whose eccentricity is just a shade less than 1. Where is its second focus? At the center of the earth!

title:0.892
EXERCISES

plain text:0.943
1. When a point P varies on an ellipse, the sum $OP + O_1P$ of its two focal distances is constant. (See Figure 6.6B.)

Gambar II.2 Contoh Hasil DLA

Menurut Shehzadi, Stricker, dan Afzal (2024), DLA dibagi menjadi dua aspek uta-

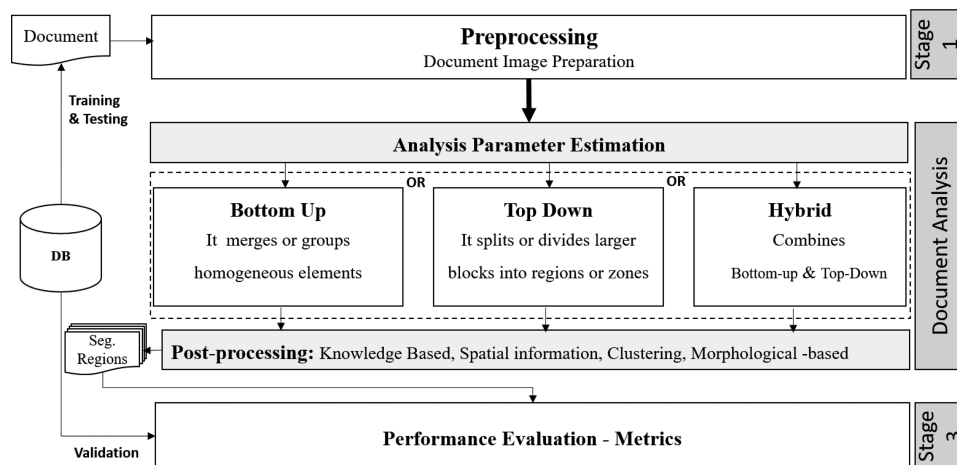
ma.

1. *Physical layout analysis*

Tahap ini berfokus pada identifikasi dan pengelompokan elemen-elemen fisik halaman secara spasial, seperti teks, gambar, dan tabel.

2. *Logical layout analysis*

Tahap ini memberikan makna semantik pada elemen-elemen tersebut, misalnya judul, paragraf, dan *header*, serta memahami hubungan hierarki dan urutan pembacaannya.



Gambar II.3 Kerangka Kerja DLA secara Umum (Binmakhashen dan Mahmoud 2019)

Mengingat tingginya keragaman tata letak dokumen, hingga kini belum ada algoritma DLA universal yang mampu mencakup semua jenis tata letak atau memenuhi seluruh tujuan analisis (Binmakhashen dan Mahmoud 2019). Oleh karena itu, Binmakhashen dan Mahmoud (2019) mengemukakan suatu kerangka kerja DLA yang bersifat umum, logis, dan sistematis. Kerangka kerja tersebut berfungsi sebagai alur dalam proses analisis tata letak dokumen dan terdiri atas lima fase utama.

1. *Preprocessing*

Fase *preprocessing* bertujuan mengubah *raw document image* menjadi gambar yang siap diproses oleh metode DLA. Langkah ini memastikan bahwa gambar masukan memenuhi prasyarat analisis, seperti telah dibersihkan, dinormalisasi, dan dikoreksi dari kemiringan (*de-skewed*). Kerusakan pada dokumen dapat berasal dari degradasi alami, seperti tinta yang menyebar atau tulisan yang memudar, maupun degradasi tambahan, seperti kerusakan alat pemindai atau kemiringan saat pemindaian. Oleh karena itu, efek negatif dari masalah tersebut perlu diminimalkan sebelum proses analisis tata letak dimulai.

2. *Analysis parameter estimation*

Parameter analisis merupakan pengukuran yang telah ditentukan sebelumnya dan digunakan untuk membantu metode DLA mengontrol proses analisis dokumen. Parameter ini dibagi menjadi dua jenis utama:

(a) *Model-driven parameters*

Parameter yang diestimasi untuk menyesuaikan model DLA agar memenuhi tujuan analisisnya.

Contoh: jumlah node atau lapisan pada *Multi-layer Perceptron* (MLP), serta penentuan *initial weights* untuk pelatihan model.

(b) *Data-driven parameters*

Parameter yang dihitung menggunakan berbagai pengukuran berdasarkan kumpulan data yang diberikan.

Contoh: rata-rata jarak antarbaris, jarak antarkata, dan rata-rata tinggi atau lebar karakter.

3. *Layout analysis*

Inti dari kerangka kerja DLA terletak pada fase ini, yaitu proses segmentasi halaman. Terdapat tiga strategi utama dalam analisis tata letak:

(a) *Bottom-Up*

Pendekatan ini memulai analisis dari bagian terkecil dalam dokumen, seperti piksel atau kelompok kecil elemen yang saling terhubung. Bagian-bagian yang mirip kemudian digabung menjadi area yang lebih besar sampai terbentuk zona dokumen yang utuh. Strategi ini merupakan cara yang paling umum digunakan dalam DLA.

(b) *Top-Down*

Analisis dimulai dari wilayah besar (misalnya, tingkat dokumen), kemudian wilayah tersebut dipecah menjadi zona yang lebih kecil (seperti kolom teks) berdasarkan aturan homogenitas. Proses berhenti ketika tidak ada lagi pemisahan zona yang dapat dilakukan.

(c) *Hybrid*

Mengintegrasikan kedua strategi sebelumnya (*bottom-up* dan *top-down*). Walaupun jarang digunakan, pendekatan ini dinilai lebih tangguh dan mampu menangani tata letak dokumen yang kompleks.

4. *Post-processing*

Fase *post-processing* merupakan langkah opsional dalam sebagian besar algoritma DLA. Tujuannya adalah untuk meningkatkan dan menggeneralisasi hasil algoritma agar dapat diterapkan pada jenis tata letak lain, sekaligus mengompensasi kekurangan hasil segmentasi agar lebih akurat.

5. *Performance evaluation*

Evaluasi kinerja DLA mencakup dua tugas utama.

(a) *Physical analysis*

Bertujuan mendeteksi struktur dokumen dan mengidentifikasi batas wilayah sejenis. Evaluasi dilakukan melalui pencocokan antara hasil segmentasi dan *ground-truth* pada tingkat piksel atau wilayah.

(b) *Logical Analysis*

Bertugas memberi label pada wilayah hasil deteksi (seperti gambar, tabel, dan paragraf).

BAB III

ANALISIS MASALAH

III.1 Analisis Kondisi Saat Ini

Transformasi digital merupakan hal yang penting bagi berbagai organisasi untuk meningkatkan efisiensi dalam penyimpanan, pemrosesan, dan penyaluran informasi (Ma dkk. 2025). Organisasi yang melaksanakan digitalisasi dokumen fisik umumnya melakukan proses pemindaian, yaitu konversi dokumen fisik menjadi representasi gambar digital yang biasanya disimpan dalam format PDF, JPEG, atau TIFF. Melalui proses ini, dokumen yang sebelumnya hanya dapat diakses secara manual kini dapat disimpan dan diarsipkan dalam bentuk digital, sehingga lebih mudah diakses serta dapat mengurangi risiko kerusakan.

Meskipun pemindaian berhasil mengubah dokumen dari bentuk fisik menjadi digital untuk tujuan pengarsipan, hasilnya masih berupa gambar semata. Output dari pemindaian standar bersifat *non-machine-readable* (Dias dan Lopes November 2023). Artinya, sistem hanya mengenali berkas hasil pemindaian sebagai kumpulan piksel tanpa memahami isi maupun struktur teks yang terkandung di dalamnya. Kondisi ini menimbulkan berbagai keterbatasan, seperti sulitnya melakukan pencarian teks, ekstraksi data, dan pengeditan. Oleh karena itu, tantangan utama dalam transformasi digital dokumen bukan hanya terletak pada perubahan format, tetapi juga pada bagaimana hasil digitalisasi tersebut dapat dimanfaatkan secara optimal sesuai dengan kebutuhan pengguna.

III.2 Analisis Kebutuhan

Berdasarkan hasil analisis kondisi saat ini yang disajikan pada Subbab III.1, diketahui bahwa proses digitalisasi yang dilakukan melalui pemindaian masih memiliki sejumlah keterbatasan, terutama dalam hal pemanfaatan hasil digitalisasi secara optimal. Oleh karena itu, diperlukan penerapan digitalisasi dengan pendekatan yang

lebih efektif melalui pengembangan sistem yang menawarkan fitur lebih beragam guna mendukung pengelolaan data digital dan mempermudah pengguna dalam menjalankan aktivitasnya.

III.2.1 Identifikasi Masalah Pengguna

Keterbatasan dari dokumen pindaian yang bersifat *non-machine-readable* secara langsung berdampak pada alur kerja pengguna. Berikut adalah identifikasi masalah utama yang dihadapi:

1. Tidak efisien dalam pencarian informasi spesifik

Pengguna tidak dapat menemukan informasi di dalam dokumen hasil pemindaian. Ketika diperlukan pencarian data spesifik, fungsi pencarian seperti Ctrl+F tidak dapat digunakan. Akibatnya, pengguna harus membaca dokumen hasil pemindaian secara manual, halaman demi halaman. Proses tersebut terbukti memakan waktu, tidak efisien, dan berpotensi menimbulkan kesalahan manusia.

2. Keterbatasan aksesibilitas

Keterbatasan aksesibilitas juga menjadi permasalahan penting. Dokumen yang hanya berupa gambar tidak dapat dipahami oleh teknologi bantu seperti pembaca layar, sehingga pengguna dengan keterbatasan penglihatan tidak dapat mengakses informasi di dalamnya. Hal ini menimbulkan hambatan digital serta menciptakan lingkungan kerja yang kurang mendukung bagi semua pengguna.

3. Keterbatasan integrasi dokumen hasil pemindaian

Data yang masih berformat gambar tidak dapat diekstraksi secara otomatis untuk diolah atau digunakan kembali dalam sistem lain. Kondisi ini membatasi organisasi dalam mengotomatisasi proses kerja yang berkaitan dengan pengelolaan dokumen. Akibatnya, informasi penting dari hasil pemindaian dokumen fisik tidak dapat dimanfaatkan secara optimal untuk mendukung kegiatan organisasi, terutama yang berkaitan dengan otomatisasi alur kerja.

III.2.2 Kebutuhan Fungsional

Keterbatasan dari dokumen pindaian yang bersifat *non-machine-readable* secara langsung berdampak pada alur kerja pengguna. Permasalahan yang telah diidentifikasi pada bagian III.2.1 perlu diselesaikan melalui pengembangan sistem baru dengan berbagai kemampuan fungsional. Kebutuhan fungsional tersebut menjelaskan fitur yang harus dimiliki sistem agar mampu mengatasi kendala tersebut dan men-

dukung otomatisasi proses kerja secara lebih efektif. Rincian kebutuhan fungsional sistem disajikan pada Tabel III.1.

Tabel III.1 Daftar Kebutuhan Fungsional Sistem

| Kode | Kebutuhan Fungsional |
|------|---|
| FR01 | Hasil digitalisasi dokumen harus disimpan dalam format yang dapat diedit dan dicari, seperti DOCX |
| FR02 | Tata letak dokumen digital yang dihasilkan harus menyerupai dokumen fisik aslinya |
| FR03 | Elemen dalam dokumen digital yang dihasilkan, seperti teks, tabel, dan gambar, harus dapat dipilih, disalin, serta dipindahkan ke aplikasi lain |

III.2.3 Kebutuhan Nonfungsional

Kebutuhan nonfungsional berfokus pada aspek kualitas dan kinerja sistem yang akan dikembangkan. Aspek ini tidak secara langsung menggambarkan fungsi utama sistem, tetapi sangat berpengaruh terhadap keandalan, efisiensi, serta pengalaman pengguna dalam pengoperasian sistem. Rincian kebutuhan nonfungsional sistem disajikan pada Tabel III.2.

Tabel III.2 Daftar Kebutuhan Nonfungsional Sistem

| Kode | Parameter | Kebutuhan Nonfungsional |
|-------|----------------------|--|
| NFR01 | <i>Accuracy</i> | Tingkat akurasi penentuan elemen harus mencapai minimal 90% untuk dokumen dengan kualitas baik dan teks yang jelas |
| NFR02 | <i>Compatibility</i> | Sistem harus mendukung format input berkas gambar, yaitu PDF, JPEG, PNG, dan TIFF |
| NFR03 | <i>Performance</i> | Sistem harus mampu memproses dokumen dengan waktu respons maksimal 1 menit per halaman |
| NFR04 | <i>Performance</i> | Sistem harus mampu melakukan ekstraksi tabel dengan akurasi minimal 80% untuk tabel dengan struktur yang jelas |
| NFR05 | <i>Reliability</i> | Jika gambar input tidak dapat diproses dengan akurat, sistem harus menampilkan pesan peringatan yang informatif |
| NFR06 | <i>Security</i> | Sistem harus melakukan validasi tipe dan ukuran file sebelum proses unggah untuk mencegah masuknya file berbahaya |
| NFR07 | <i>Usability</i> | Antarmuka sistem harus intuitif sehingga pengguna dapat mengunggah dan memproses dokumen maksimal dalam tiga langkah |
| NFR08 | <i>Usability</i> | Sistem harus memberikan indikator visual yang jelas selama proses digitalisasi, agar pengguna mengetahui status pemrosesan |

III.3 Analisis Pemilihan Solusi

III.3.1 Alternatif Solusi

Beberapa pendekatan OCR berbasis *deep learning* dapat diterapkan dalam digitalisasi dokumen fisik agar dokumen digital yang dihasilkan dapat diedit sekaligus mempertahankan tata letak aslinya. Berbeda dengan OCR konvensional yang hanya mampu mengenali karakter tanpa memahami konteks dan struktur dokumen, pendekatan OCR berbasis *deep learning* yang bisa dibuat mampu memahami konteks dan struktur dokumen. Teknologi ini menggabungkan kemampuan *text recognition* dengan pemahaman tata letak dokumen, sehingga dapat mengidentifikasi berbagai elemen seperti text, tabel, dan gambar dengan akurat. Pendekatan-pendekatan ini

menjadi solusi yang efektif bagi permasalahan digitalisasi dokumen yang kompleks, dengan penjelasan rinci sebagai berikut.

III.3.1.1 Kombinasi YOLO, CRNN, dan *Post-processing*

Model *You Only Look Once* (YOLO) digunakan untuk melakukan deteksi objek pada tingkat halaman. Dalam konteks ini, objek berarti berbagai area seperti area teks, tabel, atau gambar. YOLO membagi gambar ke dalam kisi-kisi. Kemudian, untuk setiap sel kisi, model memprediksi keberadaan elemen tertentu di dalamnya. Hasilnya adalah serangkaian kotak pembatas. Masing-masing kotak ini memiliki koordinat yang terdiri dari nilai x, y, lebar, dan tinggi yang menunjukkan letak setiap elemen di halaman. Proses ini memungkinkan sistem mengetahui batas yang tepat untuk tiap elemen dokumen.

Setelah mendapatkan koordinat, sistem akan memotong setiap elemen menjadi citra-citra terpisah. Setiap potongan citra ini kemudian diproses lebih lanjut sesuai jenisnya. Sebagai contoh, citra area teks akan diolah menggunakan model *Convolutional Recurrent Neural Network* (CRNN) yang berfungsi mengenali teks per baris. Model ini bekerja dengan membaca citra teks secara sekuensial, mengekstrak fitur spasial melalui lapisan konvolusi, lalu menerjemahkannya menjadi rangkaian karakter menggunakan lapisan rekuren dan dekode.

Untuk area tabel, sistem melakukan analisis struktur garis di dalam area hasil pemotongan. Pola horizontal dan vertikal yang berpotongan dikenali menggunakan deteksi tepi untuk menemukan batas sel. Informasi ini digunakan untuk membangun ulang tabel secara digital dengan struktur baris dan kolom yang sama. Area gambar, sebaliknya, hanya disalin langsung tanpa diubah.

Tahap akhir adalah penyusunan ulang. Sistem menggunakan koordinat setiap kotak pembatas hasil deteksi awal untuk menempatkan kembali teks, tabel, dan gambar ke posisi yang sama di halaman digital. Dengan begitu, hasil akhirnya mempertahankan tata letak asli dokumen fisik.

III.3.1.2 PaddleOCR

PaddleOCR menyediakan alur kerja lengkap yang secara otomatis menangani seluruh proses, mulai dari deteksi teks, pembacaan isi, hingga pemahaman struktur dokumen. Dalam konteks konversi dokumen fisik ke digital dengan tata letak yang sama, komponen PP-Structure menjadi bagian yang paling penting.

Langkah awal dilakukan dengan deteksi teks menggunakan model seperti DBNet yang berfungsi mendeteksi lokasi teks di dalam gambar. Model ini bekerja dengan memprediksi area yang memiliki kemungkinan mengandung teks, kemudian menghasilkan kotak pembatas poligonal yang melingkupi setiap baris atau area teks. Area non-teks seperti tabel dan gambar akan memiliki pola berbeda dan dapat dipisahkan melalui modul analisis tata letak bawaan.

Setelah area teks ditemukan, sistem melakukan pengenalan teks dengan model CRNN untuk membaca isi teksnya per baris. Sementara itu, modul khusus untuk pengenalan tabel mendeteksi pola garis horizontal dan vertikal pada area tabel. PP-Structure mampu mengonversi hasil tersebut langsung menjadi format tabel digital dengan isi setiap sel diisi berdasarkan posisi teks yang terbaca. Untuk gambar, sistem menyimpannya sebagai area media dengan posisi tetap.

Tahap berikutnya adalah pemulihan tata letak, yaitu proses menyusun ulang semua elemen, termasuk teks, tabel, dan gambar, berdasarkan koordinat asli yang tersimpan selama deteksi awal. PaddleOCR menyimpan koordinat dalam satuan piksel relatif terhadap ukuran halaman, sehingga hasil digital dapat dibangun ulang secara proporsional. Dengan pendekatan ini, posisi setiap area teks, tabel, dan gambar di dokumen digital akan identik dengan posisi pada dokumen fisik aslinya.

III.3.1.3 Kombinasi EasyOCR dan Layout Parser

Pendekatan ini mengombinasikan EasyOCR untuk membaca teks dengan Layout Parser untuk memahami struktur tata letak dokumen. Proses dimulai dengan menerapkan Layout Parser untuk melakukan analisis visual halaman. Layout Parser bekerja menggunakan model deteksi seperti Detectron2 yang telah dilatih untuk mengenali berbagai jenis elemen dokumen, misalnya teks, tabel, dan gambar.

Ketika dijalankan, Layout Parser menghasilkan kumpulan kotak pembatas untuk setiap elemen beserta label jenisnya. Sistem kemudian memotong setiap area berdasarkan koordinat tersebut. Area teks selanjutnya dikirim ke EasyOCR untuk proses pengenalan. Dalam proses ini, sistem membaca huruf demi huruf dan mengubahnya menjadi teks digital yang dapat disunting.

Untuk area tabel, sistem menggunakan pendekatan dua tahap. Layout Parser pertama mengenali bahwa area tersebut adalah tabel. Setelah itu, dilakukan analisis garis dan kisi menggunakan metode deteksi tepi serta analisis jarak antarsegmen untuk menemukan struktur baris dan kolom. Hasilnya diubah menjadi tabel digital dengan

ukuran sel yang disesuaikan berdasarkan jarak fisik antarbaris dan kolom pada gambar asli. Sementara itu, gambar hanya disalin sebagai elemen dengan posisi tetap.

Dalam tahap rekonstruksi halaman, setiap elemen digital yang telah dihasilkan akan ditempatkan kembali dengan mengacu pada koordinat awal kotak pembatas. Layout Parser memfasilitasi proses ini dengan menyediakan format posisi. Fitur ini memungkinkan sistem membangun ulang halaman digital yang mempertahankan seluruh tata letak aslinya. Dengan demikian, teks, tabel, dan gambar muncul di posisi yang sama seperti pada dokumen fisik.

III.3.2 Analisis Penentuan Solusi

Untuk memperoleh alternatif solusi yang paling sesuai untuk dikembangkan, langkah evaluasi awal adalah melakukan analisis kualitatif. Analisis ini difokuskan untuk mengidentifikasi kelebihan dan kekurangan dari setiap alternatif solusi yang diajukan.

Proses identifikasi ini penting karena berfungsi sebagai landasan pertimbangan yang berimbang. Dengan memetakan kedua sisi dari setiap opsi, potensi manfaat yang ditawarkan oleh keunggulan solusi dapat dimaksimalkan, sekaligus potensi risiko yang mungkin timbul dari kekurangannya dapat diminimalkan. Pemaparan rinci mengenai perbandingan kelebihan dan kekurangan dari seluruh alternatif solusi disajikan pada Tabel III.3 berikut.

Tabel III.3 Kelebihan dan Kekurangan Masing-Masing Alternatif Solusi

| Solusi | Kelebihan | Kekurangan |
|--|--|--|
| Kombinasi YOLO, CRNN, dan <i>Post-processing</i> | <ol style="list-style-type: none"> 1. Setiap bagian bisa diatur secara terpisah untuk hasil terbaik 2. Sangat baik dalam menentukan lokasi presisi untuk setiap elemen seperti teks, tabel, dan gambar | <ol style="list-style-type: none"> 1. Tidak ada sistem yang menyatukan semuanya. Perlu usaha besar untuk membuat deteksi, pengenalan, dan perbaikan tata letak bekerja bersamaan 2. Perlu membuat algoritma tambahan dari nol untuk mendeteksi struktur tabel, terutama untuk tabel yang rumit dan tanpa garis |
| PaddleOCR | <ol style="list-style-type: none"> 1. Menyediakan satu paket terintegrasi untuk deteksi, pengenalan teks, analisis tata letak, dan pembacaan tabel 2. Mampu membaca dan mengubah struktur tabel menjadi format digital secara otomatis | <ol style="list-style-type: none"> 1. Sulit mengubah perilaku internal bila perlu penyesuaian detail 2. Hasil spasi dan pindah baris masih memerlukan perbaikan manual |

(Lanjutan ke halaman berikutnya)

Tabel III.3 — *Lanjutan*

| Solusi | Kelebihan | Kekurangan |
|-------------------------------------|---|--|
| Kombinasi EasyOCR dan Layout Parser | <ol style="list-style-type: none"> 1. Mampu membedakan dan mengklasifikasi berbagai elemen dengan akurasi tinggi 2. Komponen terpisah memudahkan penambahan logika khusus | <ol style="list-style-type: none"> 1. Perlu usaha menggabungkan hasil keduanya 2. Tidak langsung mengekstrak struktur tabel kompleks |

Analisis kualitatif pada tabel III.3 perlu didukung oleh penilaian kuantitatif yang lebih objektif dan terukur. Untuk memenuhi kebutuhan ini, akan diterapkan metode *Weighted Scoring Model* (WSM). Metode ini menyediakan kerangka kerja yang sistematis untuk menilai dan membandingkan alternatif solusi secara numerik.

Dalam metode WSM, setiap kriteria akan diberikan bobot persentase yang menyatakan prioritas kebutuhan dalam penyelesaian masalah dengan total bobot adalah seratus persen. Pada metode WSM ini sudah ditetapkan tiga kriteria evaluasi utama. Berikut adalah definisi, alasan, dan alokasi bobot untuk setiap kriteria.

1. Efektivitas (bobot 50%)

Kriteria ini menjadi prioritas utama karena mengukur secara langsung kemampuan solusi dalam menyelesaikan akar permasalahan, terutama akurasi ekstraksi dan tata letak dokumen. Sebagai inti dari keberhasilan, kriteria ini diberikan porsi bobot terbesar yaitu lima puluh persen.

2. Implementability (bobot 30%)

Kriteria ini meninjau aspek teknis dan ketersediaan sumber daya, seperti estimasi waktu, biaya, dan kompleksitas integrasi. Kriteria ini penting untuk menentukan kelayakan dan kecepatan penyelesaian, sehingga diberikan bobot tiga puluh persen.

3. Skalabilitas (Bobot 20%)

Kriteria ini mengukur kemampuan sistem untuk berkembang di masa depan, seperti menangani volume data yang lebih besar dan kebutuhan kustomisasi baru. Ini adalah pertimbangan jangka panjang yang diberi bobot dua puluh

persen.

Penerapan proses penilaian yang telah dijabarkan, yaitu pemberian skor performa pada setiap alternatif dan kalkulasinya terhadap bobot kriteria yang telah ditetapkan, telah selesai dilakukan. Hasil perhitungan kuantitatif menggunakan metode WSM ini disajikan secara rinci pada Tabel III.4.

Tabel III.5 Daftar Kebutuhan Nonfungsional Sistem

| Kriteria | Kombinasi YOLO, CRNN, dan <i>Post-processing</i> | PaddleOCR | Kombinasi EasyOCR dan Layout Parser |
|-------------------------------|--|-----------|-------------------------------------|
| Efektivitas (50%) | 5 | 4 | 5 |
| <i>Implementability</i> (30%) | 1 | 5 | 1 |
| Skalabilitas (20%) | 5 | 2 | 5 |
| Total | 3.8 | 3.9 | 3.8 |

Pada kriteria efektivitas, kombinasi YOLO, CRNN, dan *Post-processing* dan kombinasi EasyOCR dan Layout Parser sama-sama menerima skor 5. Penilaian ini didasarkan pada fakta bahwa keduanya adalah pendekatan yang bisa disesuaikan sehingga hasilnya bisa dibuat semirip mungkin. kombinasi YOLO, CRNN, dan *Post-processing* sangat unggul dalam presisi koordinat bounding box, sementara kombinasi EasyOCR dan Layout Parser unggul dalam pemahaman struktur dan klasifikasi elemen tata letak, sehingga keduanya dinilai menawarkan tingkat akurasi tertinggi. Di sisi lain, PaddleOCR memperoleh skor 4. Meskipun sangat fungsional dan memiliki kemampuan pengenalan tabel bawaan, akurasi presisi tata letaknya secara umum, seperti penanganan spasi dan perpindahan baris masih kurang.

Selanjutnya, pada kriteria *implementability*, perbedaannya menjadi sangat jauh dan menjadi faktor penentu. PaddleOCR menerima skor 5. Alasan utamanya karena PaddleOCR merupakan sebuah solusi *all-in-one*. Sistem ini menyediakan *pipeline* yang sudah terintegrasi penuh, mulai dari deteksi hingga pengenalan tabel, sehingga secara drastis mengurangi waktu dan kompleksitas pengembangan. Sebaliknya, kombinasi YOLO, CRNN, dan *Post-processing* dan kombinasi EasyOCR dan Layout Parser keduanya menerima skor terendah 1. Kedua pendekatan ini mewakili tingkat kompleksitas implementasi yang sangat tinggi, kombinasi YOLO, CRNN, dan *Post-processing* menuntut pembangunan *pipeline* manual dari nol dan kombinasi EasyOCR dan Layout Parser memerlukan usaha integrasi untuk menggabungkan

dua sistem terpisah.

Terakhir, pada kriteria skalabilitas yang berfokus pada jangka panjang, situasinya kembali berbalik. kombinasi YOLO, CRNN, dan *Post-processing* dan kombinasi EasyOCR dan Layout Parser keduanya menerima skor 5. Keunggulan utama keduanya terletak pada sifatnya yang modular, yang memungkinkan setiap komponen disesuaikan secara independen di masa depan. Di sisi lain, PaddleOCR menerima skor rendah 2. Sifatnya yang sudah terintegrasi erat justru menjadi kelemahan di sini, membuatnya kaku dan sulit untuk dikustomisasi, yang berpotensi menjadi hambatan jika diperlukan modifikasi spesifik di kemudian hari.

Berdasarkan analisis kualitatif kelebihan dan kekurangan pada Tabel III.3, serta didukung oleh hasil perhitungan kuantitatif metode WSM pada Tabel III.4, maka PaddleOCR ditetapkan sebagai solusi terpilih. PaddleOCR unggul secara signifikan karena menawarkan keseimbangan terbaik antara fungsionalitas dan kelayakan implementasi. Kelebihannya sebagai solusi *all-in-one* yang siap pakai, lengkap dengan kemampuan pengenalan tabel bawaan, memberikan nilai tertinggi pada kriteria *implementability*. Hal ini menjadikannya pilihan yang paling optimal untuk dikembangkan dibandingkan alternatif solusi lainnya yang membutuhkan usaha integrasi dan pengembangan manual yang jauh lebih kompleks.

BAB IV

DESAIN KONSEP SOLUSI

Ilustrasikan desain konsep solusi dalam bentuk model konseptual dan penjelasan secara ringkas, beserta perbedaannya dengan sistem saat ini. Ilustrasi harus dapat dibandingkan (*before and after*). Karena masih berupa proposal, bab ini hanya berisi gambar desain konsep solusi tersebut dan penjelasan perbandingannya dengan gambar sistem yang ada saat ini (yang tergambar di awal Bab III).

BAB V

RENCANA SELANJUTNYA

Jelaskan secara detail langkah-langkah rencana selanjutnya, hal-hal yang diperlukan atau akan disiapkan, dan risiko dan mitigasinya, yang meliputi:

1. Rencana implementasi, termasuk alat dan bahan yang diperlukan, lingkungan, konfigurasi, biaya, dan sebagainya.
2. Desain pengujian dan evaluasi, misalnya metode verifikasi dan validasi.
3. Analisis risiko dan mitigasi, misalnya tindakan selanjutnya jika ada yang tidak berjalan sesuai rencana.

DAFTAR PUSTAKA

- Binmakhashen, Galal M., dan Sabri A. Mahmoud. 2019. "Document Layout Analysis: A Comprehensive Survey". *ACM Comput. Surv.* (New York, NY, USA) 52 (6). ISSN: 0360-0300. <https://doi.org/10.1145/3355610>. <https://doi.org/10.1145/3355610>.
- Borovikov, Eugene. 2014. "A survey of modern optical character recognition techniques". *ArXiv abs/1412.4183*. <https://api.semanticscholar.org/CorpusID:15162390>.
- Dahl, Christian M., Torben S. D. Johansen, Emil N. Sørensen, Christian E. Westermann, dan Simon F. Wittrock. 2021. *Applications of Machine Learning in Document Digitisation*. arXiv: 2102.03239 [cs.CV]. <https://arxiv.org/abs/2102.03239>.
- Dias, Mariana, dan Carla Teixeira Lopes. November 2023. "Optimization of Image Processing Algorithms for Character Recognition in Cultural Typewritten Documents". *Journal on Computing and Cultural Heritage* 16, no. 4 (): 1–25. ISSN: 1556-4711. <https://doi.org/10.1145/3606705>. <http://dx.doi.org/10.1145/3606705>.
- Islam, Noman, Zeeshan Islam, dan Nazia Noor. 2017. *A Survey on Optical Character Recognition System*. arXiv: 1710.05703 [cs.CV]. <https://arxiv.org/abs/1710.05703>.
- Li, Junlong, Yiheng Xu, Tengchao Lv, Lei Cui, Cha Zhang, dan Furu Wei. 2022. "DiT: Self-supervised Pre-training for Document Image Transformer". Dalam *Proceedings of the 30th ACM International Conference on Multimedia*, 3530–3539. MM '22. Lisboa, Portugal: Association for Computing Machinery. ISBN: 9781450392037. <https://doi.org/10.1145/3503161.3547911>. <https://doi.org/10.1145/3503161.3547911>.

- Ma, Zhichao, Fan Huang, Lu Zhao, Fengjun Guo, Guangtao Zhai, dan Xionghuo Min. 2025. *DocIQ: A Benchmark Dataset and Feature Fusion Network for Document Image Quality Assessment*. arXiv: 2509.17012 [cs.CV]. <https://arxiv.org/abs/2509.17012>.
- Memon, Jamshed, Maira Sami, Rizwan Khan, dan Mueen Uddin. 2020. "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)". *IEEE Access*, 1–1. <https://doi.org/10.1109/ACCESS.2020.3012542>.
- Rihanna, Astuty, Henry Eryanto, dan Terrylina Arvinta Monoarfa. 2024. "Analysis of the Implementation of Digitization at the National Archives of the Republic of Indonesia". Dalam *International Student Conference on Business, Education, Economics, Accounting, and Management (ISC-BEAM)*, vol. 2. 1. Retrieved from the ISC-BEAM Journal website. Universitas Negeri Jakarta. <https://journal.unj.ac.id/unj/index.php/isc-beam/article/view/47126>.
- Shehzadi, Tahira, Didier Stricker, dan Muhammad Zeshan Afzal. 2024. *A Hybrid Approach for Document Layout Analysis in Document images*. arXiv: 2404.17888 [cs.CV]. <https://arxiv.org/abs/2404.17888>.
- Sinha, Rasha, dan Rekha B S. 2025. *Digitization of Document and Information Extraction using OCR*. arXiv: 2506.11156 [cs.CV]. <https://arxiv.org/abs/2506.11156>.
- Tan, Kheng Leong, Chi-Hung Chi, dan Kwok-Yan Lam. 2023. "Survey on Digital Sovereignty and Identity: From Digitization to Digitalization". *ACM Comput. Surv.* (New York, NY, USA) 56 (3). issn: 0360-0300. <https://doi.org/10.1145/3616400>. <https://doi.org/10.1145/3616400>.
- Tripathi, Pankaj. 2025a. "Major Limitations of OCR Technology and How IDP Systems Overcome Them". Docsumo. Accessed: 2025-10-27. Diakses pada October 27, 2025. <https://www.docsumo.com/blog/ocr-limitations>.
- . 2025b. "The Brief History of OCR Technology". Docsumo. Accessed: 2025-10-28. Diakses pada October 28, 2025. <https://www.docsumo.com/blog/optical-character-recognition-history>.

- Wang, Jilin, Michael Krumdick, Baojia Tong, Hamima Halim, Maxim Sokolov, Vadym Barda, Delphine Vendryes, dan Chris Tanner. 2023. “A Graphical Approach to Document Layout Analysis”. Dalam *Document Analysis and Recognition – ICDAR 2023*, disunting oleh G. A. Fink, R. Jain, K. Kise, dan R. Zanibbi, 14191:53–69. Lecture Notes in Computer Science. Springer, Cham. https://doi.org/10.1007/978-3-031-41734-4_4. https://doi.org/10.1007/978-3-031-41734-4_4.
- Xu, Yiheng, Minghao Li, Lei Cui, Shaohan Huang, Furu Wei, dan Ming Zhou. 2020. “LayoutLM: Pre-training of Text and Layout for Document Image Understanding”. Dalam *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1192–1200. KDD ’20. Virtual Event, CA, USA: Association for Computing Machinery. ISBN: 9781450379984. <https://doi.org/10.1145/3394486.3403172>. <https://doi.org/10.1145/3394486.3403172>.
- Zhong, Xu, Jianbin Tang, dan Antonio Jimeno Yepes. 2019. “PubLayNet: Largest Dataset Ever for Document Layout Analysis”. Dalam *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 1015–1022. <https://doi.org/10.1109/ICDAR.2019.00166>.