Pemanfaatan *Pipeline Document Parsing* untuk Rekonstruksi Dokumen Fisik ke Format Digital dengan Mempertahankan *Layout* dan Gaya Visual

Proposal Tugas Akhir

Oleh

Habib Akhmad Al Farisi
18222029

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

Desember 2025

LEMBAR PENGESAHAN

Pemanfaatan *Pipeline Document Parsing* untuk Rekonstruksi Dokumen Fisik ke Format Digital dengan Mempertahankan *Layout* dan Gaya Visual

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 5 Desember 2025

Pembimbing

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

NIP. 197006091995121002

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I

PENDAHULUAN

I.1 Latar Belakang

Di era transformasi digital, digitalisasi dokumen menjadi kebutuhan fundamental untuk meningkatkan efisiensi operasional dan mendukung keberlanjutan. Sistem manajemen dokumen sangat penting karena pengelolaan dokumen fisik menimbulkan inefisiensi signifikan, misalnya pemborosan waktu. Karyawan dapat menghabiskan 30 hingga 40% waktu kerja mereka hanya untuk mencari berkas (Xiong May 11, 2021). Selain itu, ada risiko kehilangan atau kerusakan dokumen fisik yang tinggi. Sebaliknya, dokumen digital menawarkan keamanan yang lebih baik dan kemudahan pencarian (Fleischhacker, Kern, dan Göderle 2025).

Namun, pemindaian konvensional hanya menghasilkan gambar statis seperti JPEG, PNG, dan PDF berbasis gambar yang tidak dapat dicari maupun diekstrak datanya. Untuk mengatasi keterbatasan ini, diperlukan teknologi yang mampu mengenali isi dan struktur dokumen secara akurat (Sinha dan S 2025). *Optical Character Recognition* (OCR) adalah teknologi kunci yang mengonversi gambar dokumen menjadi teks yang dapat dicari dan diekstrak, didorong oleh kemajuan dalam *deep learning* (Li, Lee, dan Liu 2025). Selain ekstraksi teks, pemahaman terhadap *layout* dokumen juga sangat penting. *Document Layout Analysis* (DLA) berperan dalam memahami struktur dan susunan elemen visual dalam dokumen, seperti deteksi teks, tabel, dan gambar (Chen dkk. 2025). Integrasi antara OCR dan DLA inilah yang membentuk *pipeline document parsing*.

Saat ini, telah tersedia berbagai *pipeline document parsing* yang mengintegrasikan OCR dan DLA untuk menghasilkan dokumen digital yang dapat dicari dan diekstrak. Namun, meskipun *pipeline* tersebut mampu mengekstrak teks dengan baik, tantangan utama masih muncul dalam mempertahankan *layout* dan gaya visual asli dokumen. Banyak sistem cenderung hanya berfokus pada ekstraksi teks, sehingga

tata letak asli tidak dipertahankan dengan baik (Pfitzmann dkk. 2022). Akibatnya, dokumen digital seringkali memerlukan penyuntingan manual untuk memperbaiki posisi elemen dan gaya teks. Kondisi ini membuat sebagian pengguna memilih kembali ke dokumen fisik karena hasil digitalisasi tidak mencerminkan bentuk aslinya (Zabukovšek, Jordan, dan Bobek 2023).

Oleh karena itu, Tugas Akhir ini bertujuan memanfaatkan *pipeline document parsing* yang sudah ada dengan mengintegrasikannya menggunakan model atau algoritma untuk meningkatkan akurasi rekonstruksi *layout* dan gaya visual. Dengan pendekatan ini, diharapkan dokumen digital yang dihasilkan dapat mempertahankan struktur dan format visual dokumen fisik secara presisi, sehingga dapat berfungsi sebagai pengganti penuh dokumen fisik dan mengatasi keterbatasan sistem digitalisasi yang ada.

I.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, berikut adalah rumusan masalah yang akan menjadi fokus dalam Tugas Akhir ini.

1. Bagaimana memanfaatkan *pipeline document parsing* yang sudah ada untuk melakukan rekonstruksi dokumen fisik ke format digital?
2. Model atau algoritma apa yang dapat ditambahkan untuk meningkatkan akurasi rekonstruksi *layout* dan gaya visual dokumen?

I.3 Tujuan

Tujuan dari Tugas Akhir ini adalah memanfaatkan *pipeline document parsing* yang sudah ada untuk menghasilkan dokumen digital dari dokumen fisik dengan mempertahankan *layout* dan gaya visual asli. Sistem yang dikembangkan tidak hanya mengekstraksi teks secara akurat, tetapi juga merekonstruksi struktur dokumen dengan memastikan penempatan elemen dan gaya teks sesuai dengan dokumen fisik aslinya. Dengan demikian, dokumen digital yang dihasilkan dapat berfungsi sebagai pengganti penuh dokumen fisik tanpa memerlukan penyuntingan manual yang signifikan.

I.4 Batasan Masalah

Bagian ini menyajikan batasan masalah yang digunakan untuk memperjelas ruang lingkup pembahasan dalam Tugas Akhir ini.

1. Jenis dokumen yang diproses dibatasi pada dokumen fisik berbahasa Indonesia dengan struktur 1 kolom.
2. Dokumen fisik hanya berisi tabel sederhana, yaitu data yang tersusun rapi dalam baris dan kolom.
3. Akurasi rekonstruksi dokumen bergantung pada kualitas deteksi dan klasifikasi elemen dokumen yang dihasilkan oleh *pipeline document parsing* yang digunakan.
4. Ukuran dokumen fisik yang dapat diproses dibatasi pada ukuran A4.
5. Gambar dokumen yang digunakan harus memiliki metadata *dots per inch* (dpi) sebagai satuan resolusi pemindaian yang memadai agar teks dan elemen visual dapat terdeteksi dengan jelas.
6. Sistem tidak dirancang untuk mengenali atau memproses dokumen yang mengalami kerusakan berat, memiliki orientasi terbalik, buram, atau tercoret secara signifikan.
7. Format dokumen digital yang dihasilkan dibatasi pada format PDF.

I.5 Metodologi

Metodologi dalam Tugas Akhir ini mencakup lima tahapan utama, yakni eksplorasi, perancangan, implementasi, pengujian, dan evaluasi. Alur Tugas Akhir 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

   Tahap ini mencakup eksplorasi terhadap *pipeline document parsing* yang tersedia dan mampu mengintegrasikan OCR dengan DLA. Dilakukan analisis terhadap berbagai *pipeline* untuk memilih solusi yang dapat mengekstraksi teks, mendeteksi elemen dokumen, serta memberikan informasi *layout* yang diperlukan untuk rekonstruksi. Selain itu, dilakukan kajian terhadap model atau algoritma tambahan yang dapat meningkatkan akurasi rekonstruksi *layout* dan gaya visual dokumen.
2. Perancangan

   Setelah *pipeline document parsing* dipilih, disusun arsitektur sistem yang memanfaatkan keluaran *pipeline* tersebut untuk melakukan rekonstruksi dokumen. Perancangan mencakup integrasi *pipeline document parsing* yang di-

pilih dengan model atau algoritma tambahan yang diperlukan untuk meningkatkan presisi rekonstruksi.

3. Implementasi

   Pada tahap ini, rancangan sistem direalisasikan melalui pengembangan program yang memproses dokumen fisik menggunakan *pipeline document parsing* yang telah dipilih. Sistem mengekstraksi informasi teks, *layout*, dan gaya visual dari keluaran *pipeline*, kemudian merekonstruksi dokumen ke format digital dengan mempertahankan posisi elemen dan gaya teks sesuai dokumen fisik asli.

4. Pengujian

   Pengujian dilakukan dengan membandingkan dokumen digital hasil rekonstruksi terhadap dokumen fisik asli. Evaluasi mencakup akurasi *layout* dan akurasi gaya visual. Jika performa belum optimal, proses iteratif dilakukan dengan mengulang tahapan sebelumnya untuk meningkatkan kualitas rekonstruksi.

5. Evaluasi

   Evaluasi meninjau hasil pengujian secara menyeluruh untuk menilai kemampuan sistem dalam mempertahankan *layout* dan gaya visual dokumen fisik pada hasil dokumen digital. Evaluasi juga menganalisis keterbatasan sistem dan memberikan rekomendasi perbaikan. Temuan ini digunakan untuk menyimpulkan tingkat keberhasilan sistem dalam merekonstruksi dokumen secara akurat dan kesesuaiannya sebagai pengganti dokumen fisik.

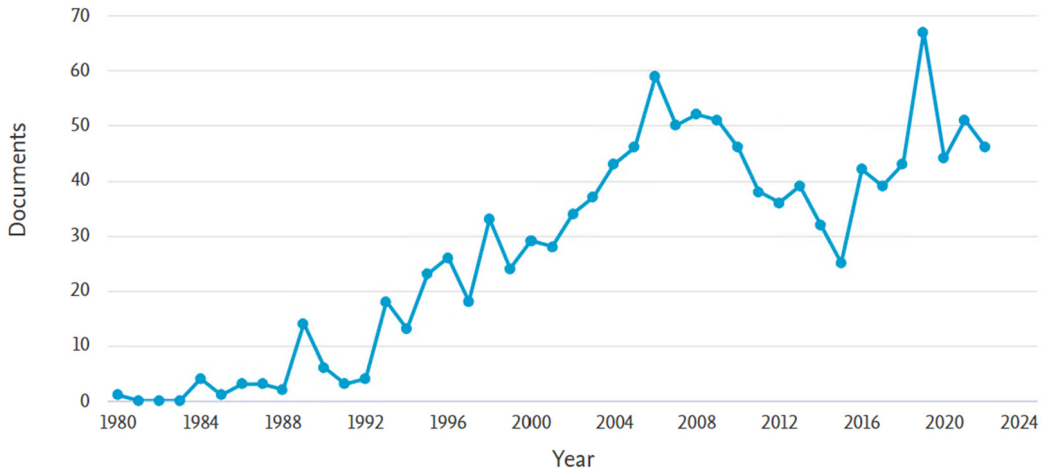
BAB II

STUDI LITERATUR

II.1 Digitalisasi Dokumen

Digitalisasi dokumen adalah proses mengubah dokumen fisik menjadi format digital sehingga lebih mudah disimpan, ditemukan, dan diakses (Zabukovšek, Jordan, dan Bobek 2023). Transformasi ini menjadi penting karena organisasi modern membutuhkan alur kerja yang cepat, efisien, dan fleksibel, terutama dalam lingkungan kerja jarak jauh dan kolaboratif.

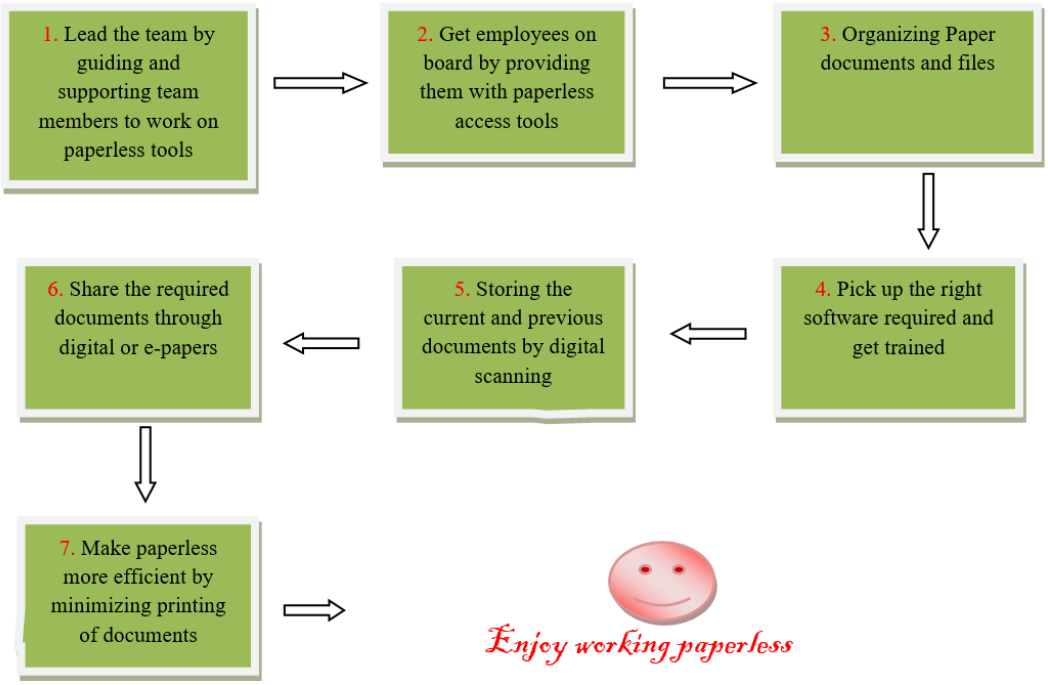
Minat terhadap digitalisasi meningkat pesat secara global. Gambar II.1 menunjukkan jumlah publikasi terkait *Document Management System* (DMS) pada basis data Scopus. Terlihat dua lonjakan besar, yaitu saat krisis finansial 2008 dan pandemi COVID-19 2020, menandakan bahwa digitalisasi menjadi solusi penting ketika proses kerja fisik terhambat (Zabukovšek, Jordan, dan Bobek 2023).



Gambar II.1 Jumlah publikasi dengan kata kunci “Document Management System” pada basis data Scopus (Zabukovšek, Jordan, dan Bobek 2023)

Keterbatasan dokumen fisik menjadi alasan utama digitalisasi dilakukan. Pencarian arsip kertas dapat menghabiskan 30–40% waktu kerja dan rentan terhadap kehilangan maupun kerusakan (Xiong May 11, 2021). Sebaliknya, dokumen digital dapat dicari dalam hitungan detik, diakses dari berbagai lokasi, serta dilindungi dengan enkripsi dan kontrol akses (Zabukovšek, Jordan, dan Bobek 2023). Digitalisasi juga mengurangi biaya penyimpanan fisik, biaya pencetakan, dan penggunaan kertas, sehingga berkontribusi pada efisiensi dan keberlanjutan lingkungan (Yousufi 2023).

Dalam implementasinya, organisasi menghadapi tantangan seperti kebutuhan investasi awal, pelatihan pengguna, dan pemilihan teknologi yang tepat. Untuk itu, Yousufi (2023) mengusulkan tujuh langkah implementasi menuju lingkungan kerja *paperless* sebagaimana ditunjukkan pada Gambar II.2. Langkah-langkah ini menekankan pentingnya perencanaan, pemilihan solusi digital, dan evaluasi berkelanjutan.



Gambar II.2 Tujuh langkah menuju lingkungan kerja *paperless* yang efisien (Yousufi 2023)

Secara keseluruhan, digitalisasi dokumen tidak hanya menggantikan media kertas, tetapi mengubah cara organisasi mengelola informasi. Dokumen digital menawarkan akses yang lebih cepat, keamanan lebih baik, keawetan jangka panjang, serta dukungan terhadap proses bisnis yang gesit dan modern (Ogilvie 2016). Oleh karena itu, digitalisasi kini menjadi pilar penting dalam transformasi digital organisasi.

II.2 *Optical Character Recognition* (OCR)

OCR merupakan perangkat lunak untuk mengonversi teks dan gambar cetak ke dalam format digital sehingga dapat digunakan oleh komputer (Islam, Islam, dan Noor 2017). Lebih spesifik, OCR adalah proses konversi citra hasil pindai dari teks cetak menjadi teks yang dapat diproses oleh mesin, baik sebagai berkas teks biasa maupun format HTML (Borovikov 2014).

II.2.1 Tahapan Utama dalam Proses OCR

Proses OCR tidak terjadi secara langsung, melainkan melalui serangkaian tahap sistematis yang dirancang untuk mengubah citra dokumen menjadi teks digital yang dapat diproses. Menurut Islam, Islam, dan Noor (2017), proses ini terbagi ke dalam enam fase utama sebagai berikut:

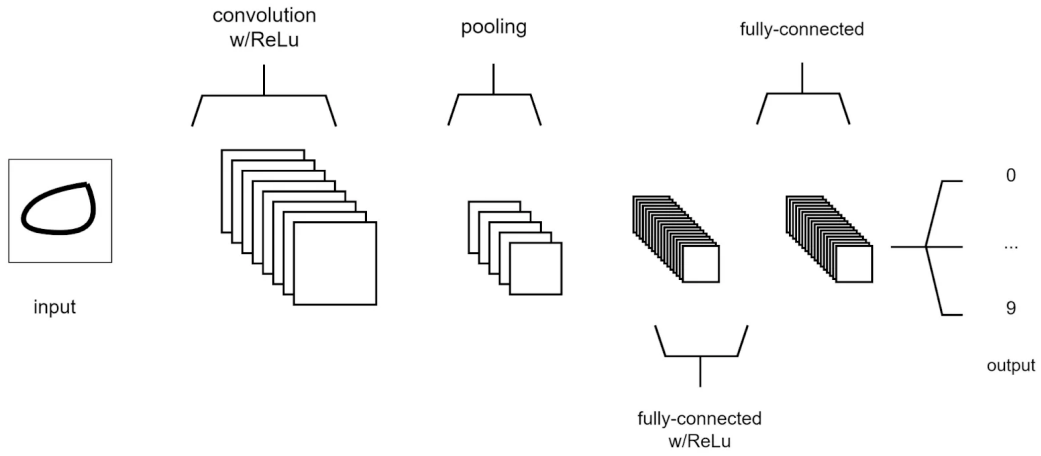
1. *Image acquisition*
   Memperoleh citra digital dan mengubahnya ke format yang dapat diproses melalui kuantisasi, binarisasi, dan kompresi.
2. *Preprocessing*
   Meningkatkan kualitas visual dengan menghilangkan *noise*, memisahkan teks dari latar belakang (*thresholding*), melakukan koreksi kemiringan (*deskew*), serta menerapkan operasi morfologi.
3. *Character segmentation*
   Memisahkan teks menjadi karakter individual menggunakan teknik seperti analisis komponen terhubung atau profil proyeksi.
4. *Feature extraction*
   Mengekstraksi fitur unik yang membedakan karakter, baik yang bersifat geometris (*loop*, *stroke*) maupun statistik (*moments*).
5. *Character classification*
   Memetakan fitur ke kategori karakter tertentu menggunakan pendekatan struktural maupun statistik, seperti Bayesian dan *neural network*.
6. *Post-processing*
   Meningkatkan akurasi keluaran melalui penggabungan beberapa *classifier*, analisis konteks, serta pemanfaatan kamus dan model probabilistik sehingga menghasilkan teks yang lebih akurat dan konsisten.

II.2.2 Perkembangan *Optical Character Recognition* Berbasis *Deep Learning*

Perkembangan OCR mengalami peningkatan signifikan setelah hadirnya pendekatan *deep learning*. OCR berbasis *deep learning* memanfaatkan *neural network* yang

mampu mempelajari pola langsung dari data dalam jumlah besar. Teknologi seperti *Convolutional Neural Network* (CNN), *Convolutional Recurrent Neural Network* (CRNN), dan *Transformer* menghasilkan peningkatan besar dalam akurasi serta kemampuan generalisasi. Model-model tersebut mampu memahami struktur visual dan konteks teks secara lebih mendalam sehingga kinerjanya melampaui metode OCR konvensional.

CNN menjadi fondasi utama dalam OCR modern. Menurut O'Shea dan Nash (2015), CNN dirancang untuk memproses data visual dengan memanfaatkan struktur alami gambar. CNN tersusun atas neuron yang memiliki tiga dimensi, yaitu tinggi, lebar, dan kedalaman. Berbeda dengan *neural network* standar, setiap neuron tidak terhubung dengan seluruh neuron pada lapisan sebelumnya. Setiap neuron hanya memproses area kecil pada gambar yang disebut *local receptive field*. Pembatasan cakupan ini membuat jumlah parameter jauh lebih sedikit sehingga proses pelatihan menjadi lebih efisien dan lebih mudah distabilkan.



Gambar II.3 Arsitektur CNN Sederhana (O'Shea dan Nash 2015)

Secara umum, CNN terdiri atas tiga komponen utama, yaitu *convolutional layer*, *pooling layer*, dan *fully-connected layer*. Ketiganya bekerja secara bertahap untuk mendeteksi pola dasar, mereduksi kompleksitas, dan menghasilkan prediksi akhir berdasarkan fitur yang telah terbentuk. Penjelasan lebih lanjut sebagai berikut:

1. *Convolutional layer*

   Lapisan ini berfungsi mendeteksi pola visual dasar seperti tepi, garis, sudut, dan lengkungan. Setiap *filter* atau *kernel* diterapkan pada seluruh area gambar melalui proses pergeseran untuk mengidentifikasi pola tertentu. Setiap kernel menghasilkan *activation map* dua dimensi yang merepresentasikan lokasi dan

intensitas pola tersebut (O'Shea dan Nash 2015). Penggunaan beberapa kernel memungkinkan model mengenali berbagai fitur visual secara bersamaan.

2. *Pooling layer*
   Lapisan ini berfungsi mereduksi ukuran *feature map*. Reduksi dilakukan melalui proses *downsampling* pada dimensi spasial, misalnya dengan *max pooling* atau *average pooling*. O'Shea dan Nash (2015) menjelaskan bahwa proses ini mengurangi jumlah parameter sekaligus meningkatkan ketahanan model terhadap variasi kecil pada gambar. Pooling membantu mempertahankan informasi penting sambil membuang detail yang kurang relevan.
3. *Fully-connected layer*
   Lapisan ini berfungsi menghasilkan prediksi kelas berdasarkan fitur yang telah diekstraksi oleh lapisan sebelumnya. Lapisan ini tersusun atas neuron yang terhubung penuh ke neuron pada dua lapisan yang berdekatan, serupa dengan struktur *neural network* standar. Lapisan ini mengubah representasi fitur menjadi keputusan akhir, misalnya memetakan citra karakter ke kelas tertentu dalam konteks OCR.

Seiring meningkatnya kebutuhan untuk mengenali teks yang berbentuk urutan, diperlukan model yang tidak hanya memahami pola visual, tetapi juga hubungan antar karakter. Kebutuhan tersebut mendorong lahirnya *Convolutional Recurrent Neural Network* (CRNN). CRNN menggabungkan kekuatan CNN untuk ekstraksi fitur visual, RNN untuk pemodelan urutan, dan *Connectionist Temporal Classification* (CTC) sebagai mekanisme transkripsi yang tidak memerlukan segmentasi karakter secara manual.

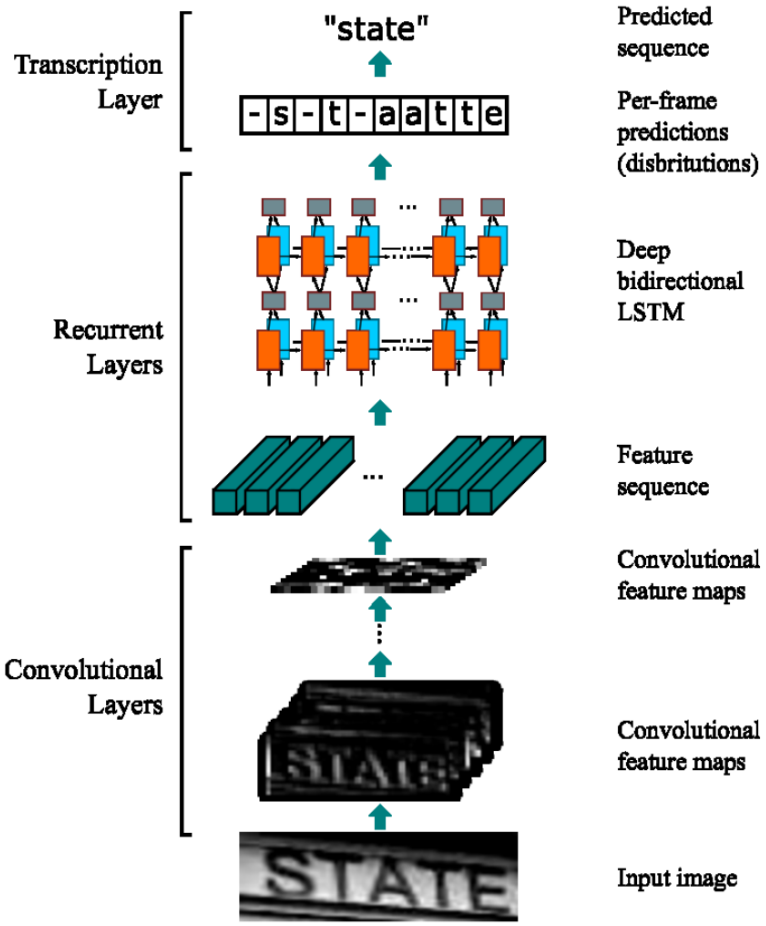
Menurut Shi, Bai, dan Yao (2015), CRNN umumnya tersusun atas tiga komponen utama, yaitu *convolutional layers*, *Bidirectional Long Short-Term Memory* (BLSTM) layers, dan *transcription layer*. Penjelasan lengkapnya sebagai berikut:

1. *Convolutional layers* Lapisan ini bertugas mengekstraksi pola visual karakter sebagaimana pada CNN. Shi, Bai, dan Yao (2015) menjelaskan bahwa lapisan konvolusi secara otomatis menghasilkan urutan vektor fitur dari gambar *input*. Setiap vektor fitur diperoleh dari kolom *feature map* yang diekstraksi secara berurutan dari kiri ke kanan. Proses ini mengubah citra dua dimensi menjadi rangkaian fitur satu dimensi yang siap diproses oleh lapisan rekuren.
2. BLSTM
   BLSTM memproses urutan fitur dari dua arah sekaligus, yaitu dari kiri ke kanan dan dari kanan ke kiri. Huang, Xu, dan Yu (2015) menjelaskan bahwa pendekatan dua arah efektif dalam menangkap ketergantungan jangka pan-

jang, terutama ketika karakter memiliki bentuk mirip atau ketika konteks dari masa depan membantu memperjelas prediksi. BLSTM memanfaatkan struktur LSTM yang mampu menyimpan informasi penting melalui mekanisme *input gate*, *forget gate*, dan *output gate*. Mekanisme ini membuat model mampu mempelajari ketergantungan jangka pendek maupun jangka panjang pada data berurutan.

3. CTC

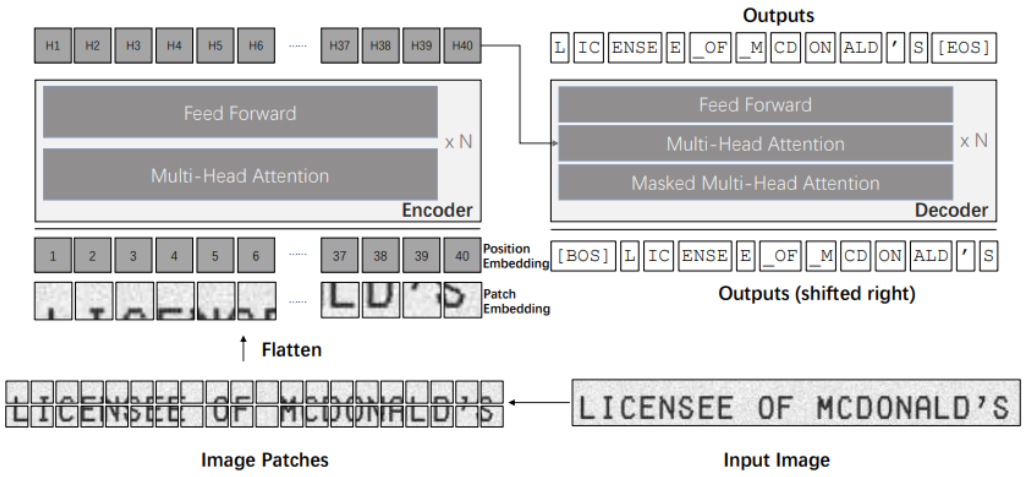
   Lapisan transkripsi ini menghasilkan urutan karakter tanpa memerlukan segmentasi eksplisit. CTC mendefinisikan probabilitas kondisional untuk urutan label berdasarkan prediksi per-frame dan mengabaikan posisi pasti setiap label (Shi, Bai, dan Yao 2015). Dengan demikian, model dapat dilatih hanya dengan pasangan gambar dan teks utuh tanpa menunjukkan lokasi setiap karakter. Shi, Bai, dan Yao (2015) menyatakan bahwa fungsi objektif pada CRNN menggunakan *negative log-likelihood* dari probabilitas urutan tersebut sehingga kebutuhan anotasi menjadi jauh lebih sederhana dan efisien.



Gambar II.4 Arsitektur CRNN (Shi, Bai, dan Yao 2015)

Menurut Shi, Bai, dan Yao (2015), CRNN mampu mengenali teks dengan panjang bervariasi tanpa perlu menentukan panjang keluaran sejak awal. CRNN dapat menerima gambar dengan dimensi berbeda dan menghasilkan prediksi dengan panjang yang menyesuaikan kontennya. Kemampuan ini sangat penting karena teks di dunia nyata memiliki panjang yang sangat beragam, mulai dari kata tunggal hingga kalimat panjang. Ghojogh dan Ghodsi (2023) menambahkan bahwa LSTM, yang menjadi komponen inti CRNN, dirancang khusus untuk mempelajari ketergantungan jangka pendek dan jangka panjang. LSTM mencapai hal tersebut melalui mekanisme gerbang yang mengatur informasi mana yang perlu disimpan, diperbarui, atau dibuang sepanjang tahapan pemrosesan urutan.

Fase selanjutnya dalam pengembangan OCR ditandai oleh pemanfaatan arsitektur *Transformer*. Arsitektur ini mulai mengambil alih peran CNN dan RNN dengan memperkenalkan mekanisme *self-attention*. Menurut Li dkk. (2022), *Transformer* mampu memodelkan hubungan global antarbagian gambar dengan memecah gambar menjadi potongan kecil (*patches*), mengubahnya menjadi vektor, serta menambahkan *positional embedding* untuk mempertahankan informasi posisi. Pemrosesan berbasis *self-attention* ini memberikan pemahaman konteks visual yang lebih menyeluruh, terutama pada *scene text* yang memiliki latar belakang kompleks.



Gambar II.5 Arsitektur OCR Berbasis *Transformer* (Li dkk. 2022)

Vaswani dkk. (2023) menjelaskan bahwa *self-attention* bekerja dengan memetakan *query* dan pasangan *key-value* menjadi keluaran yang dihitung sebagai penjumlahan berbobot dari *values*. Mekanisme ini memungkinkan model menghubungkan elemen-elemen input tanpa batasan jarak atau ketergantungan sekuensial seperti pada RNN. Islam (2023) menegaskan bahwa Vision Transformers (ViT) mampu me-

nangkap ketergantungan jangka panjang dalam konteks global melalui mekanisme *self-attention*, memberikan representasi visual yang tidak dimiliki arsitektur konvolusional tradisional.

Model OCR berbasis *Transformer* juga memperoleh keunggulan melalui integrasi *pretrained vision models* seperti ViT serta *pretrained language models*. Li dkk. (2022) menjelaskan bahwa TrOCR menggabungkan *image Transformer* dan *text Transformer* yang telah dilatih pada data tak berlabel berskala besar, sehingga keseluruhan proses pengenalan teks dapat dilakukan secara *end-to-end* tanpa memerlukan CTC atau *language model* tambahan.

Dosovitskiy dkk. (2021) menyatakan bahwa ViT tidak memiliki *image-specific inductive biases* seperti translational invariance atau lokalitas seperti pada CNN. Ketiadaan bias ini membuat *Transformer* lebih fleksibel dalam menangani variasi bentuk teks. Li dkk. (2022) juga menekankan bahwa TrOCR menghilangkan ketergantungan terhadap jaringan konvolusional dan tidak memperkenalkan *inductive bias* khusus gambar, sehingga model lebih mudah diimplementasikan, lebih sederhana dipelihara, dan mampu menangani *printed text*, *handwritten text*, maupun *scene text*.

II.3 *Document Layout Analysis* (DLA)

DLA merupakan tahapan awal yang sangat penting dalam sistem pemahaman dokumen (Binmakhashen dan Mahmoud 2019). DLA berfungsi mendeteksi dan memberikan anotasi pada struktur fisik dokumen, serta memahami pengaturan elemen-elemen di dalamnya (Shehzadi, Stricker, dan Afzal 2024). Tujuan utamanya adalah memudahkan proses analisis selanjutnya dengan mengidentifikasi blok-blok dokumen yang memiliki keseragaman serta menentukan hubungan antarblok tersebut (Binmakhashen dan Mahmoud 2019). Dengan kemampuannya mengubah dokumen tidak terstruktur menjadi format terstruktur, DLA berperan sebagai komponen utama dalam *document parsing* untuk mendukung proses identifikasi dan ekstraksi data (Shehzadi, Stricker, dan Afzal 2024).

II.3.1 Aspek Utama dalam DLA

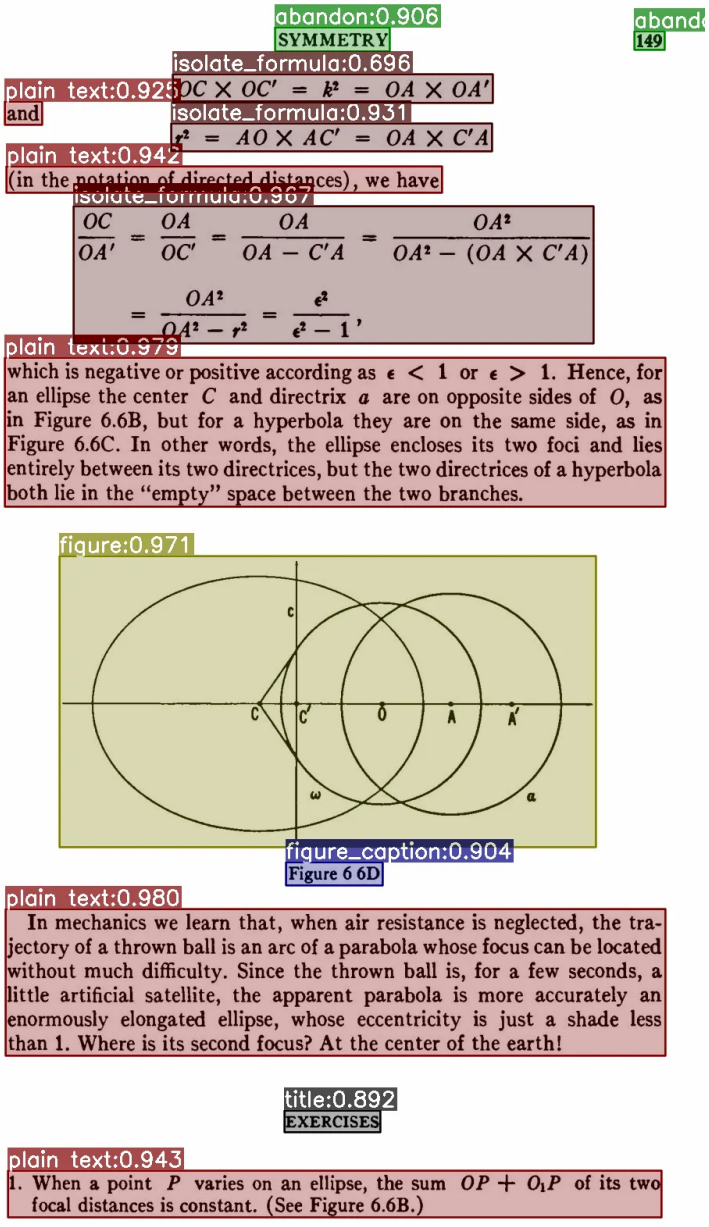
Menurut Shehzadi, Stricker, dan Afzal (2024), DLA terbagi menjadi dua aspek pokok:

1. *Physical layout analysis*
   Berfokus pada identifikasi dan pengelompokan elemen-elemen fisik halaman

secara spasial, seperti teks, gambar, dan tabel.

2. *Logical layout analysis*
   Memberikan makna semantik pada elemen-elemen tersebut, misalnya judul, paragraf, dan *header*, serta memahami hubungan hierarki dan urutan pembacaannya.

Gambar II.6 menunjukkan contoh hasil DLA pada sebuah halaman dokumen. Terlihat bahwa dokumen telah tersegmentasi menjadi beberapa wilayah yang diberi label sesuai jenis kontennya. Segmentasi seperti ini memungkinkan sistem memahami struktur dokumen secara hierarkis untuk pemrosesan lebih lanjut.

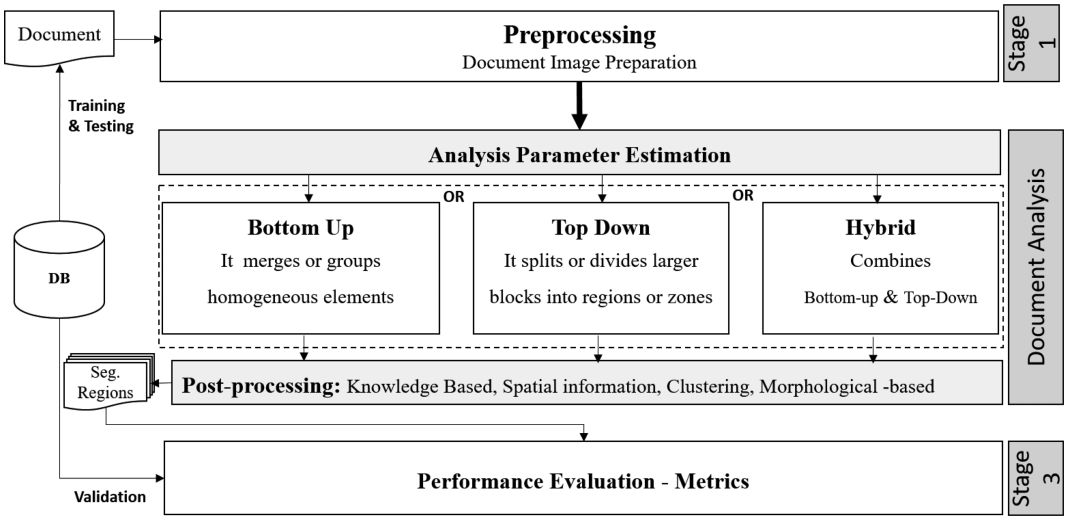


Gambar II.6 Contoh Hasil DLA

II.3.2 Kerangka Kerja Umum DLA

Menurut Binmakhashen dan Mahmoud (2019), terdapat lima fase pokok dalam kerangka kerja DLA, yaitu:

1. *Preprocessing*
   Fase ini mengubah gambar dokumen mentah menjadi gambar yang siap diproses. Prosesnya meliputi pembersihan gambar, binerisasi, dan koreksi kemiringan. Tujuannya untuk mengurangi efek negatif akibat kerusakan alami maupun teknis.
2. *Analysis parameter estimation*
   Fase ini menentukan parameter penting untuk mengendalikan proses DLA. Parameter terbagi menjadi:
   (a) *Model-driven parameters*, misalnya jumlah *node* pada ANN.
   (b) *Data-driven parameters*, misalnya rata-rata jarak antarbaris, jarak antarkata, atau tinggi karakter.
3. *Layout analysis*
   Ini merupakan inti dari DLA, yaitu segmentasi halaman menjadi wilayah bermakna. Terdapat tiga strategi pokok:
   (a) *Bottom-up*: Dimulai dari elemen kecil seperti piksel atau komponen kecil, lalu digabungkan menjadi zona besar.
   (b) *Top-down*: Dimulai dari zona besar, lalu dipecah menjadi zona kecil berdasarkan hubungan atau keseragaman.
   (c) *Hybrid*: Kombinasi kedua strategi untuk menangani tata letak kompleks.
4. *Post-processing*
   Fase opsional untuk meningkatkan dan menggeneralisasi hasil segmentasi agar lebih akurat dan kompatibel terhadap berbagai jenis dokumen.
5. *Performance evaluation*
   Fase evaluasi yang mencakup:
   (a) *Physical analysis*, yaitu membandingkan hasil segmentasi dengan *ground-truth*.
   (b) *Logical analysis*, yaitu memberi label semantik seperti paragraf, tabel, atau gambar.



Gambar II.7 Kerangka Kerja DLA secara Umum (Binmakhashen dan Mahmoud 2019)

II.3.3 Perkembangan DLA

Seiring perkembangan teknologi, metode yang digunakan dalam DLA mengalami perubahan signifikan, mulai dari pendekatan berbasis aturan hingga arsitektur *deep learning*. Menurut (Shehzadi, Stricker, dan Afzal 2024), evolusi ini dapat dibagi menjadi tiga era utama. Setiap tahap tidak hanya meningkatkan akurasi deteksi elemen dokumen, tetapi juga memperluas kemampuan sistem dalam menangani variasi *layout* yang semakin kompleks. Adapun ketiga tahap perkembangan tersebut adalah sebagai berikut:

1. *Heuristic Rule-Based* DLA

   Pada tahap awal, DLA didominasi oleh metode *heuristic rule-based* yang digunakan sebelum berkembangnya *deep learning*.

   (a) Metode *bottom-up* mengelompokkan piksel atau komponen kecil melalui proses seperti *clustering* untuk membentuk area yang seragam. Pendekatan ini mampu menangani *layout* yang rumit, namun membutuhkan komputasi tinggi.

   (b) Metode *top-down* memecah citra dokumen secara bertahap hingga terbentuk wilayah-wilayah homogen. Metode ini lebih cepat, tetapi kurang fleksibel dan hanya optimal untuk jenis dokumen tertentu.

   (c) Metode *hybrid* menggabungkan kelebihan *bottom-up* dan *top-down* sehingga lebih seimbang dalam kecepatan dan akurasi.

   Meskipun bermanfaat pada dokumen sederhana, metode heuristik cenderung kurang adaptif terhadap variasi *layout* yang kompleks dan semakin jarang di-

gunakan setelah hadirnya *deep learning*.

2. *Deep Learning-Based* DLA

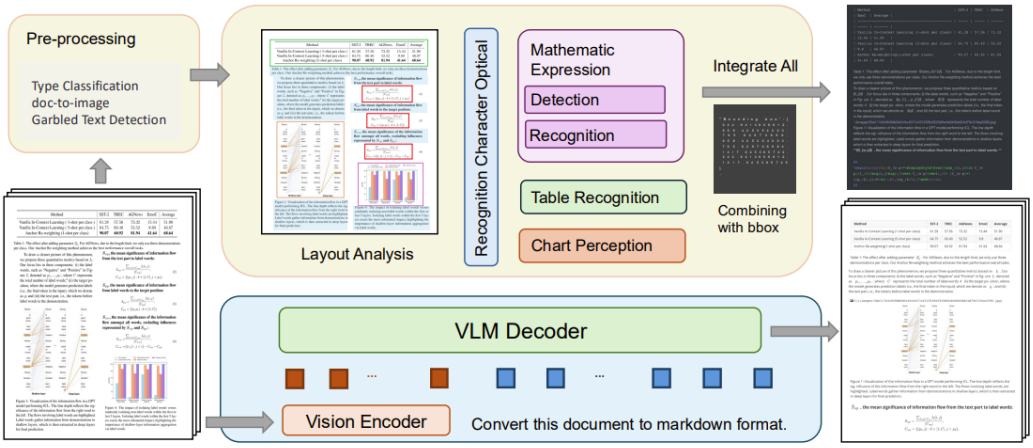
   Munculnya *deep learning* membawa peningkatan besar terhadap akurasi dan efisiensi analisis dokumen. Model seperti *Faster R-CNN* membuka peluang untuk deteksi objek pada dokumen, sedangkan *Mask R-CNN* menjadi tolok ukur baru dalam segmentasi *layout*, terutama pada dokumen dengan struktur padat seperti koran. *RetinaNet* turut digunakan untuk mendeteksi kata atau teks tertentu. Pada analisis tabel, *DeepDeSRT* memanfaatkan transformasi citra dan *fully convolutional network* dengan mekanisme *skip pooling* guna memahami struktur tabel secara lebih detail. Secara keseluruhan, pendekatan CNN meningkatkan kemampuan sistem dalam memproses *layout* dokumen yang beragam.

3. *Transformer-Based* DLA

   Perkembangan terbaru ditandai oleh pemanfaatan arsitektur *Transformer* yang mengandalkan mekanisme *self-attention* dan *positional embedding* untuk memahami konteks global dokumen. Tidak seperti CNN yang fokus pada fitur lokal, *Transformer* dapat mengintegrasikan informasi visual, tekstual, dan tata letak secara terpadu. Model seperti *DiT* meningkatkan kemampuan klasifikasi dan deteksi tabel melalui pelatihan *self-supervised*. Arsitektur *encoder–decoder* seperti *TILT* memproses ketiga modalitas secara bersamaan, sementara *LayoutLMv3* memperkuat representasi dokumen melalui pembelajaran multimodal. Namun, beberapa model seperti *DINO* masih menghadapi tantangan dalam mendeteksi objek kecil seperti judul halaman, *header*, atau *footer*. Penelitian terbaru berfokus pada peningkatan mekanisme *query* dan strategi *matching* untuk mengatasi keterbatasan tersebut.

II.4 *Document Parsing*

Menurut Zhang dkk. (2025), *document parsing* adalah teknologi esensial untuk mengonversi dokumen tidak terstruktur dan semi-terstruktur, seperti kontrak, makalah akademik, dan faktur, menjadi data terstruktur yang dapat dibaca mesin. Teknologi ini juga dikenal sebagai *document content extraction*. Tujuannya adalah mengekstrak elemen-elemen seperti teks, persamaan matematis, tabel, dan gambar dari berbagai *input* sambil mempertahankan hubungan strukturalnya. Konten yang diekstrak kemudian ditransformasikan ke dalam format terstruktur, seperti Markdown atau JSON, yang memfasilitasi integrasi ke dalam alur kerja modern.



Gambar II.8 Pendekatan Utama pada *Document Parsing* (Zhang dkk. 2025)

Zhang dkk. (2025) mengidentifikasi dua pendekatan utama dalam *document parsing* yang digunakan untuk mencapai konversi dari dokumen tidak terstruktur menjadi data terstruktur, seperti diilustrasikan pada Gambar II.8, yaitu:

1. *Modular pipeline system*
   Sistem ini melibatkan proses bertahap dan terpisah, serta terdiri atas beberapa komponen utama:
   (a) *Layout analysis*: Mendeteksi elemen struktural dokumen, seperti blok teks, paragraf, judul, gambar, tabel, dan ekspresi matematis, beserta koordinat spasial dan urutan bacanya.
   (b) *Content extraction*:
      i. *Text extraction*: Menggunakan OCR untuk mengonversi teks dalam gambar dokumen menjadi teks yang dapat dibaca mesin.
      ii. *Mathematical expression extraction*: Mendeteksi dan mengonversi simbol serta struktur matematis ke format standar, seperti LaTeX atau MathML.
      iii. *Table data and structure extraction*: Mengenali struktur tabel dan menggabungkan data yang diekstrak dengan hasil OCR.
      iv. *Chart recognition*: Mengidentifikasi jenis diagram dan mengekstrak data serta hubungan struktural.
   (c) *Relation integration*: Menggabungkan elemen-elemen yang diekstrak ke dalam struktur terpadu menggunakan koordinat spasial dari deteksi tata letak.
2. *End-to-End Approaches* dengan *Vision Language Models* (VLMs)
   Kemajuan terkini dalam *multimodal large models*, khususnya VLMs, me-

nawarkan alternatif yang menjanjikan. Model seperti GPT-4 dan QwenVL memproses data visual dan tekstual secara bersamaan, memungkinkan konversi *end-to-end* dari gambar dokumen menjadi *output* terstruktur. Meskipun demikian, pendekatan ini memiliki keterbatasan untuk tugas rekonstruksi dokumen yang presisi. VLMs menghasilkan keluaran dalam format seperti Markdown (Nougat) atau JSON (GOT) yang berfokus pada representasi semantik dan pemahaman hierarki dokumen, bukan pada penyediaan koordinat spasial seperti yang dilakukan oleh *modular pipeline system*. Selain itu, VLMs cenderung tidak menunjukkan kinerja yang konsisten melebihi *modular pipeline system* dalam tugas-tugas, seperti membedakan elemen-elemen halaman. Oleh karena itu, pendekatan ini kurang optimal untuk tujuan rekonstruksi tata letak dokumen agar identik dengan sumber aslinya.

II.5 Font Identification

Menentukan gaya huruf yang sesuai untuk gambar dokumen merupakan tantangan tersendiri, karena dalam satu gambar dokumen sangat mungkin terdapat variasi gaya huruf, misalnya antara bagian judul dan paragraf. Karena penting untuk menghasilkan output yang benar-benar menyerupai dokumen fisiknya, penentuan gaya huruf tidak dapat menggunakan metode *font generation* atau *font synthesis*. Oleh karena itu, diperlukan pendekatan *font identification*.

Salah satu pendekatan *font identification* adalah *DeepFont* yang dilatih menggunakan 2383 kategori huruf. *DeepFont* merupakan sistem berbasis *deep learning* yang dirancang untuk mengenali jenis font dari gambar teks. Sistem ini dikembangkan karena proses mengenali font secara manual sangat sulit, terutama ketika gambar berkualitas rendah, bercampur dengan latar belakang, atau mengandung distorsi. Selain itu, jumlah font sangat banyak dan sering kali memiliki perbedaan kecil yang sulit dibedakan secara visual. Untuk keperluan pelatihan, *DeepFont* menggunakan dua jenis data, yaitu data sintetis (teks yang dirender secara otomatis) dan data nyata dari foto dunia nyata. Karena data nyata sangat terbatas, *DeepFont* menggabungkan kedua jenis data tersebut agar model dapat belajar secara lebih efektif (Wang dkk. 2015).

DeepFont menggunakan arsitektur *CNN* yang terdiri dari dua bagian. Bagian pertama adalah *unsupervised cross-domain sub-network* yang dilatih menggunakan *Stacked Convolutional Auto-Encoder* (SCAE). Bagian ini bertugas mempelajari ciri visual dasar, seperti bentuk garis dan tepi huruf, baik dari data sintetis maupun data

nyata tanpa menggunakan label. Dengan pendekatan ini, model dapat memahami pola umum pada kedua sumber data sehingga mengurangi kesenjangan antara gambar sintetis dan gambar nyata. Bagian kedua adalah *supervised domain-specific sub-network* yang dilatih menggunakan data sintetis berlabel untuk mempelajari pola yang lebih kompleks sehingga mampu membedakan satu font dari font lainnya (Wang dkk. 2015).

Untuk meningkatkan performa, *DeepFont* menerapkan teknik *text data augmentation*, seperti mengubah jarak antarhuruf dan *aspect ratio*. Teknik ini membuat data sintetis tampak lebih realistis karena dalam desain grafis, huruf sering dimodifikasi untuk kebutuhan visual. Selain itu, *DeepFont* menggunakan metode pengujian *multi-scale* dan *multi-view* dengan memotong beberapa bagian gambar dan memprosesnya dalam berbagai ukuran. Hasil dari potongan-potongan tersebut kemudian digabungkan untuk mendapatkan prediksi yang lebih akurat. Dengan pendekatan ini, *DeepFont* mampu mencapai akurasi tinggi meskipun gambar memiliki kualitas yang tidak ideal (Wang dkk. 2015).

Selain menghasilkan prediksi font, *DeepFont* juga dapat menghitung kemiripan antarfont menggunakan fitur pada lapisan tertentu. Kemampuan ini memungkinkan sistem memberikan rekomendasi font yang mirip dengan font pada gambar input, yang sangat bermanfaat bagi desainer dalam mencari alternatif. *DeepFont* juga mendukung kompresi model sehingga ukurannya dapat diperkecil tanpa penurunan akurasi yang signifikan, membuatnya dapat digunakan pada perangkat dengan keterbatasan memori seperti aplikasi *mobile* (Wang dkk. 2015).

II.6 PPstructure-v3 (PaddleOCR)

PP-StructureV3 merupakan sistem *pipeline* multi-model yang dikembangkan untuk melakukan *parsing* dokumen berbasis gambar. Sistem ini mampu mengonversi gambar dokumen maupun berkas PDF menjadi keluaran terstruktur dalam format JSON dan Markdown secara akurat dan efisien (Cui dkk. 2025).

II.6.1 Arsitektur Sistem

PP-StructureV3 terdiri atas lima modul utama yang berfungsi secara berurutan untuk memproses dokumen secara menyeluruh (Cui dkk. 2025). Setiap modul memiliki peran spesifik dalam peningkatan kualitas masukan, ekstraksi informasi, serta analisis struktur dokumen. Gambaran lengkap kelima modul tersebut dapat dilihat pada Gambar II.9.

1. *Preprocessing*

   Modul ini menyiapkan gambar dokumen sebelum memasuki tahap pemrosesan lanjutan. Dua komponen utama yang digunakan ialah model klasifikasi orientasi berbasis PP-LCNet dan model *text image unwarping* berbasis UVDoc untuk memperbaiki distorsi geometris. Dengan desain tersebut, modul ini efektif dalam menangani gambar dokumen berkualitas rendah yang mengalami rotasi ataupun distorsi sehingga menghasilkan masukan yang lebih bersih untuk tahap berikutnya.
2. OCR

   Modul OCR memanfaatkan PP-OCRv5 dengan *preprocessing* yang dinonaktifkan untuk mendeteksi serta mengenali seluruh teks dalam gambar dokumen. Dibandingkan versi sebelumnya, PP-OCRv5 memberikan peningkatan signifikan pada pengenalan layout vertikal, tulisan tangan, serta karakter Mandarin yang jarang muncul. Peningkatan ini penting untuk menjaga akurasi ekstraksi teks pada berbagai jenis dokumen.
3. *Layout analysis*

   Modul analisis tata letak terdiri atas dua model yang saling melengkapi. Pertama, PP-DocLayout-plus yang merupakan versi optimal dari PP-DocLayout dan dirancang untuk mendeteksi layout kompleks seperti majalah dan koran multi-kolom, laporan dengan banyak tabel, dokumen ujian, tulisan tangan, serta layout berorientasi vertikal. Kedua, model *region detection* yang mengidentifikasi artikel berbeda dalam satu halaman, misalnya satu halaman koran yang berisi beberapa artikel, sehingga tiap elemen dapat diasosiasikan dengan artikel yang benar dan urutan pembacaan dapat dipulihkan secara tepat.
4. *Document items recognition*

   Berdasarkan prediksi hasil *layout detection*, sistem mengenali isi setiap elemen halaman menggunakan metode yang sesuai. Empat jenis elemen utama yang diproses dalam tahap ini adalah sebagai berikut.

   (a) Tabel

   Sistem menggunakan PP-TableMagic, sebuah sistem lengkap yang mencakup klasifikasi orientasi tabel, klasifikasi tipe bingkai, deteksi sel berbasis *object detection*, serta pengenalan struktur untuk menghasilkan keluaran dalam format HTML.

   (b) Formula

   Pengenalan formula dilakukan menggunakan PP-FormulaNet_plus, versi peningkatan dari PP-FormulaNet yang mampu mengenali potongan gambar formula dan menghasilkan kode LaTeX secara akurat. Pening-

katan mencakup panjang token hingga 2560, perluasan dataset formula kompleks, serta dukungan terhadap formula yang mengandung karakter Mandarin.

(c) Grafik

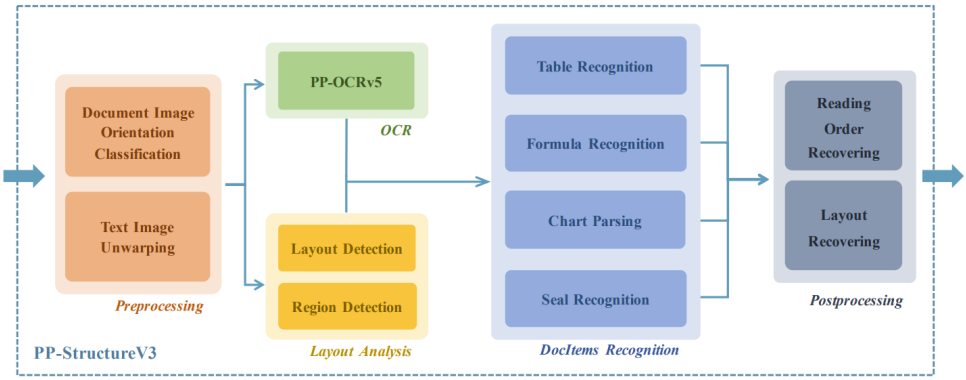
Sistem memakai PP-Chart2Table, yaitu model *vision–language end-to-end* yang ringan untuk mengekstraksi data dari berbagai jenis grafik seperti histogram, *line chart*, dan *pie chart*, lalu mengonversinya menjadi tabel berformat Markdown. Model ini ditingkatkan melalui mekanisme *Shuffled Chart Data Retrieval*, penataan ulang token yang cermat, *data synthesis pipeline* berbasis RAG, serta distilasi LLM dua tahap dengan data *out-of-distribution*.

(d) Stempel

Untuk pengenalan stempel, sistem menggunakan PP-OCRv4_seal yang dirancang untuk mengenali berbagai bentuk stempel, termasuk oval dan lingkaran. Sistem ini dilengkapi model deteksi teks melengkung yang mampu meluruskan teks yang bengkok serta model pengenalan teks generik untuk meningkatkan akurasi.

5. *Post-processing*

Setelah seluruh elemen dikenali, modul ini menyusun kembali hubungan antar komponen dalam dokumen, seperti menghubungkan gambar atau tabel dengan keterangan (*caption*) yang sesuai serta memulihkan urutan pembacaan. Modul ini menggunakan versi peningkatan dari algoritma X–Y Cut yang memberikan hasil lebih baik pada layout kompleks seperti majalah, koran, dan dokumen dengan *typeset* vertikal.



Gambar II.9 *Pipeline* PP-StructureV3 (Zabukovšek, Jordan, dan Bobek 2023)

II.6.2 Keunggulan Utama

PP-StructureV3 mencapai performa *state-of-the-art* dalam *document parsing* untuk bahasa Inggris dan Mandarin. Meskipun hanya memiliki sekitar 0,091 miliar parameter, sekitar 5/1000 dari Qwen2.5-VL-72B, model ini mampu menghasilkan kinerja yang setara atau bahkan melampaui model berparameter miliaran (Cui dkk. 2025).

Selain efisien, PP-StructureV3 juga komprehensif dalam menangani berbagai elemen dokumen, mulai dari teks, tabel, formula, grafik, hingga stempel. Kemampuannya dalam memproses layout kompleks seperti dokumen multi-kolom, halaman koran dengan banyak artikel, serta orientasi vertikal menjadikannya sangat adaptif untuk berbagai kebutuhan digitalisasi dokumen. Keluaran berupa JSON dan Markdown yang terstruktur memungkinkan pemanfaatan langsung untuk berbagai aplikasi lanjutan tanpa memerlukan pemrosesan tambahan yang signifikan (Cui dkk. 2025).

II.7 DocLayout-YOLO

DocLayout-YOLO merupakan algoritma yang dirancang khusus untuk menganalisis struktur dan tata letak berbagai jenis dokumen secara otomatis. Tujuannya adalah mengatasi tantangan klasik dalam analisis dokumen, yaitu mempertahankan keseimbangan antara kecepatan pemrosesan dan akurasi hasil. Berbeda dengan metode-metode sebelumnya yang menggunakan pendekatan *multimodal* (menggabungkan fitur visual dan teks) yang cenderung lebih lambat meskipun akurat, atau metode *unimodal* (hanya menggunakan fitur visual) yang lebih cepat namun kurang presisi, DocLayout-YOLO berhasil mencapai keunggulan pada kedua aspek tersebut melalui optimalisasi khusus pada tahap *pre-training* dan arsitektur model (Zhao dkk. 2024).

Berdasarkan pengembangnya Zhao dkk. (2024), berikut merupakan kelebihan atau inovasi yang dimiliki DocLayout-YOLO.

1. Kumpulan Data *Pre-training* DocSynth-300K
   Inovasi DocLayout-YOLO yang pertama adalah penciptaan kumpulan data *pre-training* berskala besar bernama DocSynth-300K, yang terdiri dari 300 ribu dokumen sintetis dengan keragaman tinggi. Data ini dihasilkan menggunakan algoritma *Mesh-candidate BestFit*, sebuah metode yang memandang penyusunan dokumen sebagai permasalahan *2D packing problem*. Algoritma tersebut bekerja dengan cara mengambil berbagai komponen dasar dokumen

seperti teks dalam berbagai ukuran huruf, tabel dengan beragam bentuk, gambar, dan elemen-elemen lainnya yang kemudian menyusunnya secara optimal pada halaman dokumen sambil mempertimbangkan prinsip-prinsip desain seperti *alignment*, *density*, dan estetika visual. Proses ini memastikan bahwa setiap dokumen sintetis yang dihasilkan tidak hanya beragam dalam hal jenis elemen, tetapi juga dalam hal tata letak keseluruhan, mencakup format satu kolom, dua kolom, banyak kolom, hingga gaya makalah akademis, majalah, dan koran. Proses *pre-training* pada DocSynth-300K memberikan fondasi yang kuat bagi model untuk memahami berbagai jenis dokumen sebelum dilatih lebih lanjut pada tugas-tugas spesifik.

(a) Elemen yang disintesis

Meliputi berbagai hierarki judul, *main body text*, *ignored text* seperti *header* dan *footer*, gambar dan *caption*, tabel beserta keterangan dan *footer*nya, serta rumus matematika yang berdiri sendiri dan keterangannya.

(b) Augmentasi data

Proses ini mensimulasikan berbagai kondisi nyata melalui augmentasi data yang mencakup *random flip*, penyesuaian kecerahan dan kontras, *random crop*, ekstraksi tepi menggunakan filter Sobel, serta *elastic transformation* dan penambahan *Gaussian noise*.

2. Modul *Global-to-Local Controllable Receptive Module* (GL-CRM)

Inovasi kedua adalah pengembangan modul *Global-to-Local Controllable Receptive Module* (GL-CRM), sebuah komponen arsitektur *neural network* yang dirancang khusus untuk menangani variasi skala elemen dokumen secara efektif, mulai dari elemen kecil seperti judul satu baris hingga elemen besar seperti tabel yang memenuhi seluruh halaman. Modul ini bekerja dengan mengekstrak dan mengintegrasikan fitur-fitur pada berbagai tingkat skala dan granularitas melalui lapisan konvolusi yang berbagi bobot dengan penerapan tingkat dilasi yang bervariasi. Fitur-fitur yang telah digabungkan selanjutnya diproses menggunakan lapisan konvolusi ringan untuk menghasilkan *saliency mask* yang berperan sebagai *gate* dalam menentukan *relative importance weight*. Keluaran akhir diperoleh melalui *output projector* ringan yang memanfaatkan *shortcut connection* untuk menggabungkan fitur terintegrasi dengan *initial features*, sehingga informasi penting tetap terjaga selama proses transformasi. GL-CRM mengatasi tantangan variasi skala tersebut melalui pendekatan hierarkis tiga tingkat:

(a) *Global*: Menggunakan *kernel* konvolusi besar dengan dilasi yang diperluas untuk menangkap pola skala halaman.

(b) Blok: Menggunakan *kernel* yang lebih kecil untuk mendeteksi sub-blok.
(c) Lokal: Memproses informasi semantik detail dengan modul ringan.

BAB III

ANALISIS MASALAH

III.1 Analisis Kondisi Digitalisasi Dokumen Saat Ini

Transformasi digital merupakan hal 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. Keluaran dari pemindaian standar bersifat *non-machine-readable* (Dias dan Lopes 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 penyuntingan. Oleh karena itu, tantangan utama dalam transformasi digital dokumen tidak hanya terletak pada perubahan format, tetapi juga pada cara hasil digitalisasi tersebut dapat dimanfaatkan secara optimal sesuai dengan kebutuhan pengguna.

III.2 Analisis Kebutuhan Digitalisasi Dokumen

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 Subbab 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

mendukung 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 bersifat *searchable*, seperti PDF
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	*Accuracy*	Tingkat akurasi penentuan elemen harus mencapai minimal 75% untuk dokumen dengan kualitas baik dan teks yang jelas
NFR02	*Compatibility*	Sistem harus mendukung format input berkas gambar hasil pemindaian, yaitu PDF berbasis gambar, JPEG, dan PNG
NFR03	*Performance*	Sistem harus mampu memproses dokumen dengan waktu respons maksimal 1,5 menit per halaman
NFR04	*Reliability*	Jika gambar input tidak dapat diproses, sistem harus menghasilkan log error yang informatif dan mengembalikan kode status kesalahan yang sesuai
NFR05	*Security*	Sistem harus melakukan validasi tipe dan ukuran berkas sebelum pemrosesan untuk mencegah masuknya berkas berbahaya atau berkas yang tidak sesuai spesifikasi

III.3 Analisis Pemilihan Solusi

Analisis pemilihan solusi dilakukan untuk menentukan pendekatan yang paling tepat dalam mentransformasi dokumen fisik menjadi dokumen digital yang menyerupai bentuk aslinya. Proses ini mempertimbangkan kebutuhan sistem, karakteristik

data dokumen, serta kemampuan masing-masing pendekatan dalam menjalankan tahapan utama seperti deteksi tata letak, pengenalan teks, dan rekonstruksi struktur visual. Setiap alternatif dinilai berdasarkan ketepatan hasil, stabilitas performa, serta tingkat fleksibilitasnya ketika diterapkan pada variasi dokumen. Dengan mempertimbangkan faktor-faktor tersebut, pemilihan solusi dapat dilakukan secara lebih terarah dan sesuai dengan tujuan digitalisasi, yaitu menghasilkan representasi digital yang akurat secara tekstual sekaligus konsisten secara visual.

III.3.1 Alternatif Solusi

Dalam memproses gambar dokumen agar dapat direkonstruksi menjadi dokumen digital yang menyerupai bentuk fisiknya, tersedia berbagai *pipeline document parsing* yang dapat dimanfaatkan untuk memperoleh setiap elemen yang terdapat pada dokumen. *Pipeline* tersebut umumnya sudah mencakup tahap-tahap penting seperti deteksi tata letak, segmentasi elemen, hingga ekstraksi informasi, sehingga tiap elemen yang teridentifikasi dapat diproses lebih lanjut secara sistematis. Pemanfaatan *pipeline* yang telah tersedia menjadi pilihan yang efisien karena dapat menghemat waktu pengembangan serta mengurangi kebutuhan sumber daya komputasi. Meskipun demikian, baik *pipeline* siap pakai maupun *pipeline* yang dibangun secara mandiri tetap memerlukan tahap *layout reconstruction* untuk memastikan hasil akhir merepresentasikan dokumen asli secara utuh.

Berdasarkan hasil perbandingan pada Tabel III.3 dan Tabel III.4, DocLayout-YOLO dipilih sebagai model *document layout analysis* karena memberikan akurasi deteksi elemen yang unggul dan stabil pada dokumen cetak. Untuk tahap OCR, model *Transformer OCR* (TrOCR) dipilih karena menunjukkan kinerja terbaik dalam pengenalan teks dengan variasi tipografi. Dengan demikian, *pipeline* yang dibangun secara mandiri menggabungkan DocLayout-YOLO pada tahap deteksi elemen dokumen dan TrOCR pada tahap pengenalan teks. Hasil deteksi dari DocLayout-YOLO menyediakan koordinat serta klasifikasi blok dokumen yang kemudian diekstraksi sebagai input bagi TrOCR pada proses pengenalan teks per elemen.

Tahap *layout reconstruction* kemudian digunakan untuk menyusun kembali urutan baca, menggabungkan hasil pengenalan dari setiap elemen, serta merekonstruksi struktur dokumen secara menyeluruh. Pada tahap ini, dilakukan pula penyesuaian visual yang diperlukan agar representasi dokumen tetap konsisten dengan bentuk aslinya. Penyesuaian tersebut mencakup normalisasi ukuran elemen, penyelarasan tata letak, serta estimasi atribut tipografi seperti jenis huruf dan ukuran huruf berdasarkan informasi geometris dari *bounding box* dan resolusi citra. Kombinasi proses

tersebut memastikan bahwa hasil akhir tidak hanya akurat secara tekstual, tetapi juga selaras secara visual dengan struktur dan format dokumen asli.

Tabel III.3 Perbandingan Model *Layout Analysis* Berdasarkan Rata-Rata Performa (Ouyang dkk. 2025)

Model	**Backbone**	**Params**	**Average**
DiT-L	ViT-L	361,6M	26,90
LayoutLMv3	RoBERTa-B	138,4M	28,84
DocLayout-YOLO	v10m	19,6M	**47,38**
SwinDocSegmenter	Swin-L	223M	35,89
GraphKD	R101	44,5M	27,10
DOCX-Chain	-	-	21,27

Tabel III.4 Hasil Evaluasi (*Word-level Recall*, *Precision*, dan F1) pada *Dataset* SROIE (Li dkk. 2022)

Model	**Recall**	**Precision**	**F1**
CRNN	28,71	48,58	36,09
Tesseract OCR	57,50	51,93	54,57
H&H Lab	96,35	96,52	96,43
MSOLab	94,77	94,88	94,82
CLOVA OCR	94,30	94,88	94,59
TrOCR-LARGE	96,59	96,57	**96,58**

III.3.1.1 Kombinasi DocLayout-YOLO, TrOCR, dan *Layout Reconstruction*

Model DocLayout-YOLO digunakan untuk melakukan deteksi elemen pada tingkat halaman. Dalam konteks DLA, elemen tersebut mencakup area teks, tabel, maupun gambar. DocLayout-YOLO membagi citra halaman ke dalam kisi-kisi dan memanfaatkan arsitektur GL-CRM untuk mengakomodasi variasi skala elemen dokumen. Untuk setiap sel kisi, model memprediksi keberadaan elemen tertentu dengan memanfaatkan fitur global, blok, dan lokal secara hierarkis. Hasil proses ini berupa sejumlah *bounding box* yang masing-masing memiliki koordinat pusat, lebar, dan tinggi, sehingga lokasi setiap elemen pada halaman dapat diidentifikasi secara akurat, termasuk untuk elemen berukuran kecil maupun sangat besar.

Koordinat hasil deteksi tersebut kemudian digunakan untuk memotong setiap elemen menjadi citra terpisah. Setiap potongan diproses sesuai jenis elemennya. Khusus area teks, pengenalan karakter dilakukan menggunakan model TrOCR. Berbeda dari pendekatan sekuensial seperti CRNN, TrOCR memanfaatkan arsitektur *Vision*

Transformer pada tahap enkoder untuk mengekstraksi representasi visual citra teks, kemudian mendekodenya menjadi rangkaian karakter melalui *Transformer decoder*. Pendekatan ini memberikan akurasi yang lebih stabil untuk variasi tipografi, ukuran huruf, serta kondisi citra dokumen yang tidak seragam.

Tahap terakhir adalah *layout reconstruction*, yaitu menyusun kembali seluruh elemen ke dalam posisi yang konsisten dengan dokumen asli. Penyusunan dilakukan dengan memanfaatkan koordinat *bounding box* dari DocLayout-YOLO untuk menata kembali elemen secara geometris, serta mengintegrasikan hasil pengenalan teks dari TrOCR. Selain itu, dilakukan pula estimasi atribut tipografi berdasarkan informasi dari *bounding box* dan resolusi citra. Penyesuaian visual tersebut mencakup normalisasi ukuran elemen dan penyelarasan tata letak agar representasi dokumen tetap konsisten dengan bentuk aslinya. Kombinasi proses tersebut memastikan bahwa hasil akhir tidak hanya akurat secara tekstual, tetapi juga selaras secara visual dengan struktur dan format dokumen asli.

III.3.1.2 Model *Pipeline Document Parsing*

Berbagai model *pipeline document parsing* telah dikembangkan untuk meningkatkan kualitas pemrosesan dokumen. Pada penelitian yang dilakukan oleh Wei, Sun, dan Li (2025), dilakukan evaluasi terhadap enam model *pipeline* dan diperoleh bahwa PP-StructureV3 memberikan kinerja yang lebih unggul dibandingkan lima model lainnya. Tabel III.5 menyajikan hasil penilaian terhadap keenam model tersebut. Semakin rendah *overall score* yang diperoleh, semakin baik kinerja model *pipeline* yang dinilai.

Tabel III.5 Perbandingan Model-Model *Pipeline Document Parsing* (Wei, Sun, dan Li 2025)

Model	Text	Formula	Table	Order	Overall
Dolphin	0,352	0,465	0,258	0,35	0,356
Marker	0,085	0,374	0,609	0,116	0,296
Mathpix	0,105	0,306	0,243	0,108	0,191
MonkeyOCR-1.2B	0,062	0,295	0,164	0,094	0,154
PPstructure-v3	0,073	0,295	0,162	0,077	**0,152**

Model PP-Structure bekerja sebagai sebuah *pipeline* berurutan untuk memahami isi dokumen secara menyeluruh. Proses dimulai dari *layout detection* yang mengidentifikasi blok-blok penting dalam halaman, seperti paragraf teks, tabel, gambar, atau rumus. Setelah struktur halaman terbagi menjadi elemen-elemen tersebut, se-

tiap elemen diproses menggunakan modul khusus. Sebagai contoh, teks diekstraksi menggunakan OCR dan tabel dikenali strukturnya melalui *table structure recognition*. Hasil dari setiap modul kemudian digabungkan kembali dalam urutan yang sesuai dengan tata letak dokumen asli.

Setelah seluruh elemen dikenali, *pipeline* dilanjutkan dengan tahap *layout reconstruction* untuk menghasilkan dokumen akhir yang rapi serta mudah digunakan. Pada tahap ini, hasil ekstraksi teks, struktur tabel, posisi gambar, dan elemen lain disusun kembali agar menyerupai dokumen fisik aslinya, baik dari segi urutan, kolom, maupun proporsi tata letak. Proses rekonstruksi juga mencakup estimasi atribut tipografi berdasarkan informasi geometris dari hasil deteksi elemen, sehingga memastikan konsistensi tampilan dokumen hasil digitalisasi dengan dokumen asli. Sistem kemudian menggabungkan seluruh elemen tersebut ke dalam keluaran digital yang bersifat *searchable*, seperti PDF, sehingga pengguna dapat menyalin teks, memilih tabel, atau mengubahnya menjadi format DOCX sehingga dapat dilakukan penyuntingan terhadap dokumen tersebut. Tahap rekonstruksi inilah yang memastikan konsistensi visual dan fungsionalitas dokumen hasil digitalisasi.

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.6.

Tabel III.6 Kelebihan dan Kekurangan Masing-Masing Alternatif Solusi

Solusi	**Kelebihan**	**Kekurangan**
Kombinasi DocLayout-YOLO, TrOCR, dan *Layout Reconstruction*	1. Akurasi deteksi tinggi untuk elemen bervariasi 2. Kontrol penuh atas setiap tahap	1. Tidak memahami konteks dokumen
PP-StructureV3 dan *Layout Reconstruction*	1. Kinerja PP-StructureV3 terbukti unggul dibandingkan model *pipeline* lainnya 2. Pemahaman struktur dokumen secara menyeluruh	1. Kurang fleksibel untuk modifikasi 2. Kurang kontrol atas detail proses

Analisis kualitatif pada Tabel III.6 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 100% dan setiap kriteria memiliki skala 1–10. Pada metode WSM ini sudah ditetapkan tiga kriteria evaluasi utama. Berikut adalah definisi, alasan, dan alokasi bobot untuk setiap kriteria.

1. Akurasi digitalisasi (bobot 50%)
   Kriteria ini menjadi prioritas utama karena mengukur seberapa akurat sistem dapat mengonversi dokumen fisik menjadi dokumen digital, mencakup akurasi pengenalan teks, deteksi elemen, dan rekonstruksi tata letak visual. Sebagai inti dari keberhasilan digitalisasi dokumen, kriteria ini diberikan porsi bobot terbesar yaitu lima puluh persen.
2. Fidelitas visual (bobot 30%)
   Kriteria ini menilai kemampuan sistem untuk mempertahankan tampilan dokumen asli, termasuk tata letak, tipografi, dan struktur visual dalam dokumen digital hasil konversi. Kriteria ini penting untuk memastikan dokumen digital tetap konsisten dan dapat digunakan sebagaimana dokumen fisik aslinya, sehingga diberikan bobot tiga puluh persen.
3. *Implementability* (bobot 20%)
   Kriteria ini meninjau aspek teknis dan ketersediaan sumber daya, seperti es-

timasi waktu, kompleksitas integrasi, dan kemudahan implementasi. Kriteria ini penting untuk menentukan kelayakan dan kecepatan penyelesaian sehingga diberikan 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.7.

Tabel III.7 Perbandingan Skor Solusi Berdasarkan Kriteria Penilaian

Kriteria	**Bobot**	**DocLayout-YOLO + TrOCR**	**PP-StructureV3**
Akurasi Digitalisasi	50%	7,5	8,5
Fidelitas Visual	30%	6,0	8,5
Implementability	20%	6,0	8,0
Skor Total	**100%**	**6,9**	**8,4**

Pada kriteria akurasi digitalisasi, solusi kombinasi DocLayout-YOLO, TrOCR, dan *layout reconstruction* memperoleh skor 7,5. DocLayout-YOLO mampu mendeteksi elemen dokumen dengan akurasi tinggi melalui fitur hierarkis GL-CRM, sementara TrOCR memberikan pengenalan teks yang baik. Namun, karena setiap komponen bekerja terpisah dan tidak memahami konteks dokumen, proses *layout reconstruction* membutuhkan logika heuristik yang kompleks dan rentan menghasilkan kesalahan urutan baca, terutama pada tata letak multikolom atau elemen tumpang tindih. Hal ini juga berdampak pada fidelitas visual yang hanya mencapai skor 6,0, sebab estimasi atribut tipografi masih sangat bergantung pada aturan manual sehingga konsistensinya sulit dijaga.

Sebaliknya, PP-StructureV3 dengan *layout reconstruction* terintegrasi memperoleh skor 8,5 untuk akurasi digitalisasi dan 8,5 untuk fidelitas visual. Model ini memahami struktur dokumen secara holistik dan mampu menghasilkan urutan baca, hubungan spasial, serta atribut tipografi secara otomatis dan konsisten. *Pipeline* yang matang membuat proses digitalisasi lebih stabil dan keluaran dokumen dapat langsung digunakan dalam bentuk PDF yang *searchable* atau berkas DOCX yang dapat diedit. Keterbatasannya hanya muncul pada tata letak yang sangat tidak umum, tetapi dampaknya relatif kecil.

Dari sisi *implementability*, kombinasi DocLayout-YOLO, TrOCR, dan *layout reconstruction* memperoleh skor 6,0 karena membutuhkan integrasi manual yang kom-

pleks dan rentan propagasi galat, sedangkan PP-StructureV3 memperoleh skor 8,0 berkat *pipeline* terintegrasi yang meminimalkan kebutuhan pengembangan logika tambahan. Berdasarkan bobot penilaian dan hasil metode WSM, PP-StructureV3 dengan *layout reconstruction* menjadi solusi terpilih dengan skor total 8,4, melampaui kombinasi DocLayout-YOLO, TrOCR, dan *layout reconstruction* yang memperoleh 6,9. Solusi PP-StructureV3 menawarkan keseimbangan terbaik antara akurasi digitalisasi, fidelitas visual, dan kemudahan implementasi, sehingga paling sesuai untuk kebutuhan digitalisasi dokumen fisik.

Berdasarkan analisis kualitatif kelebihan dan kekurangan pada Tabel III.6 serta didukung oleh hasil perhitungan kuantitatif metode WSM pada Tabel III.7, solusi PP-StructureV3 dan *layout reconstruction* ditetapkan sebagai solusi terpilih. Solusi PP-StructureV3 unggul secara signifikan karena menawarkan keseimbangan terbaik untuk digitalisasi dokumen fisik menjadi dokumen digital dengan skor total 8,4 dibandingkan solusi kombinasi DocLayout-YOLO, TrOCR, dan *layout reconstruction* yang memperoleh skor 6,9. Keunggulan utama solusi PP-StructureV3 terletak pada kriteria akurasi digitalisasi dengan skor 8,5 dan fidelitas visual dengan skor 8,5, berkat *pipeline* terintegrasi yang sudah matang dan kemampuan pemahaman struktur dokumen secara holistik. Kelebihan solusi PP-StructureV3 sebagai sistem *end-to-end* yang siap pakai dengan keluaran terstruktur otomatis, dilengkapi tahap *layout reconstruction* yang mencakup penyusunan ulang elemen dan estimasi atribut tipografi, tanpa memerlukan pengembangan logika rekonstruksi manual yang kompleks, menjadikannya pilihan yang paling optimal untuk dikembangkan. Solusi PP-StructureV3 mampu menghasilkan dokumen digital yang tidak hanya akurat secara tekstual, tetapi juga mempertahankan fidelitas visual dokumen fisik aslinya dengan sangat baik, sehingga sangat sesuai untuk kebutuhan digitalisasi dokumen.

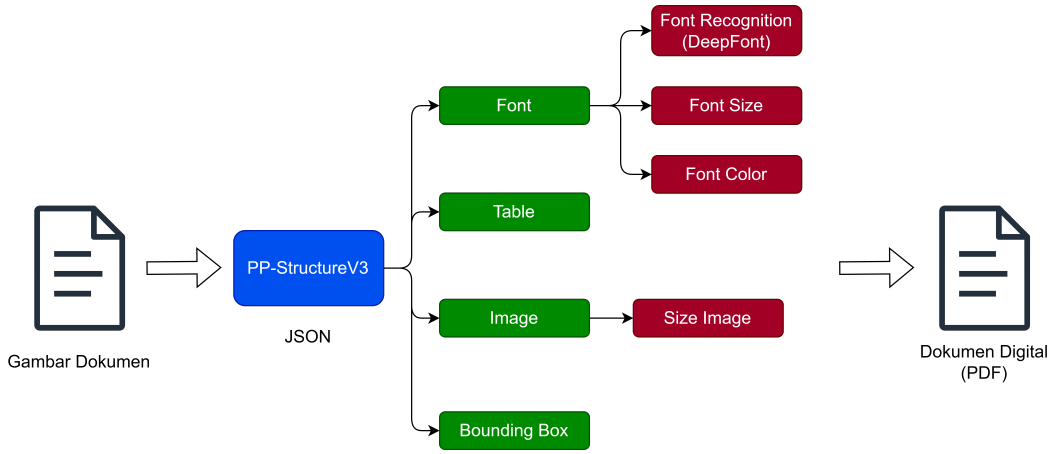
BAB IV

DESAIN KONSEP SOLUSI

Desain konsep solusi disusun berdasarkan alur pemrosesan mulai dari masukan berupa citra dokumen hingga keluaran berupa dokumen digital yang telah direkonstruksi. Keseluruhan alur sistem yang diusulkan mencakup ekstraksi teks, deteksi elemen dokumen, analisis *layout*, dan rekonstruksi gaya visual. Tahapan-tahapan ini saling terhubung untuk memastikan bahwa informasi struktur dan tampilan dokumen dipertahankan secara akurat, sehingga hasilnya adalah dokumen digital dalam format PDF yang tidak hanya dapat dicari dan diekstrak isinya, tetapi juga memiliki struktur visual yang serupa dengan dokumen fisik asli. Ilustrasi lengkap proses tersebut ditunjukkan pada Gambar IV.1 dengan penjabaran sebagai berikut:

1. *Document acquisition*
   Dokumen fisik dipindai untuk menghasilkan citra digital beresolusi memadai sebagai masukan sistem.
2. *Processing* oleh *pipeline document parsing*
   Citra kemudian diproses untuk memperoleh informasi dasar dokumen, meliputi:
   (a) Hasil *OCR* berupa teks yang dapat dicari.
   (b) Deteksi elemen seperti paragraf, tabel, dan gambar.
   (c) Analisis *layout* untuk menentukan posisi dan struktur setiap elemen.
3. *Post-processing* deteksi dan struktur
   Keluaran *pipeline* disempurnakan melalui langkah-langkah seperti penyesuaian *bounding box* dan pemisahan elemen.
4. Ekstraksi dan penetapan *visual style*
   Informasi visual tambahan, seperti *font family*, *font size*, *line spacing*, serta karakteristik elemen lainnya, dipetakan untuk mendukung proses rekonstruksi.
5. Rekonstruksi dokumen digital
   Semua informasi yang telah disusun diproyeksikan ke *digital canvas*. Sistem

menempatkan elemen sesuai posisinya dan menerapkan gaya visual agar hasil akhir menyerupai dokumen fisik.



Gambar IV.1 Desain Konsep Solusi

BAB V

RENCANA SELANJUTNYA

V.1 Rencana Implementasi

Implementasi sistem dilakukan dengan memanfaatkan perangkat lunak dan pustaka pendukung yang diperlukan untuk pemrosesan dokumen, pelatihan model, serta rekonstruksi tata letak. Rencana implementasi mencakup spesifikasi perangkat yang digunakan, lingkungan pengembangan, estimasi biaya yang diperlukan, serta linimasa pengerjaan sistem.

V.1.1 Perangkat dan Pustaka

Perangkat lunak dan pustaka yang digunakan dalam pengembangan sistem ini mencakup lingkungan komputasi, bahasa pemrograman, *pipeline* pemrosesan, dan model pembelajaran mesin. Daftar lengkap perangkat yang digunakan beserta fungsinya ditunjukkan pada Tabel V.1.

Tabel V.1 Daftar Perangkat yang Digunakan

Kategori	**Perangkat**	**Kegunaan**
Lingkungan komputasi	Google Colab Pro	Eksekusi model, *GPU runtime*
Bahasa pemrograman	Python	Implementasi *pipeline*
Pipeline	PP-StructureV3	*Pipeline* untuk memproses gambar dokumen
Model	DeepFont	Model untuk memperoleh informasi *font family*

V.1.2 Lingkungan Implementasi

Pengembangan sistem dilakukan menggunakan kombinasi perangkat lokal dan komputasi awan untuk mengoptimalkan efisiensi pengembangan. Perangkat lokal berupa laptop digunakan untuk pengembangan kode dan pengujian awal, sedangkan Google Colab Pro digunakan untuk pelatihan model yang membutuhkan akselerasi GPU. Spesifikasi laptop yang digunakan ditunjukkan pada Tabel V.2.

Tabel V.2 Spesifikasi Laptop untuk Lingkungan Pengembangan

Komponen	**Spesifikasi**
Tipe Laptop	Acer Swift SF314
Prosesor	AMD Ryzen 5 3500U, 4-core, 2,1 GHz
RAM	6 GB
Penyimpanan	SSD 512 GB
Sistem Operasi	Windows 11 Home 64-bit

Mengingat keterbatasan spesifikasi laptop, khususnya kapasitas RAM yang terbatas dan GPU terintegrasi, digunakan layanan Google Colab Pro untuk proses komputasi intensif. Layanan ini menyediakan akses GPU yang lebih *powerful* dan durasi sesi yang lebih panjang dibandingkan versi gratis, sehingga cocok untuk pelatihan model pembelajaran mesin yang membutuhkan sumber daya komputasi tinggi.

V.1.3 Estimasi Biaya

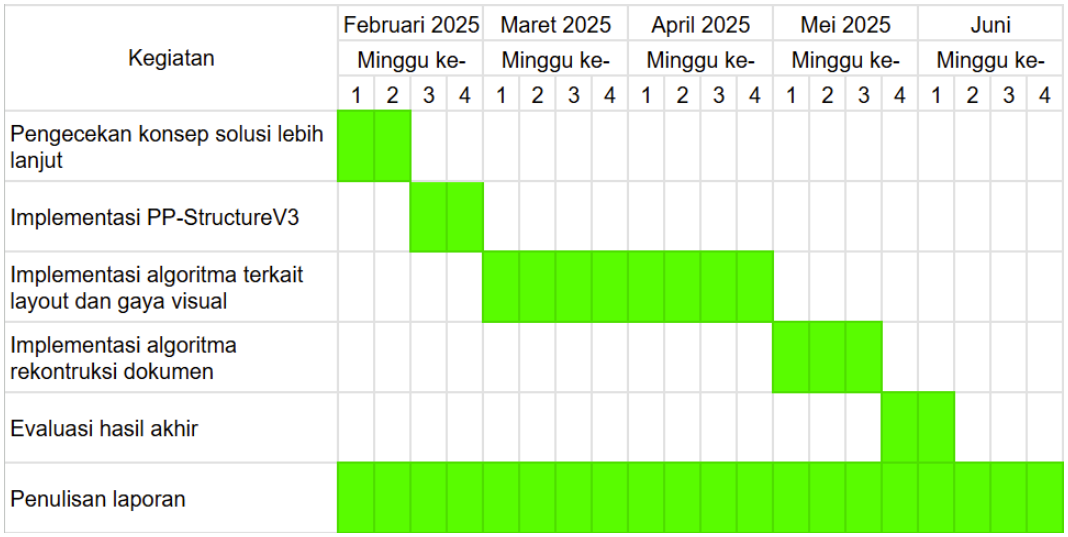
Biaya pengembangan sistem berasal dari langganan layanan komputasi awan yang diperlukan untuk pelatihan model selama periode pengembangan. Estimasi biaya untuk pengembangan sistem selama tiga bulan ditunjukkan pada Tabel V.3.

Tabel V.3 Estimasi Biaya Pengembangan

Item	**Keterangan**	**Biaya**
Google Colab Pro	Langganan bulanan (4 bulan)	Rp800.000,00
Total		**Rp800.000,00**

V.1.4 Linimasa Pengerjaan

Pengerjaan sistem dilakukan secara bertahap sesuai dengan linimasa yang telah disusun untuk memastikan setiap tahapan dapat diselesaikan tepat waktu. Linimasa pengerjaan sistem secara keseluruhan ditunjukkan pada Gambar V.1.



Gambar V.1 Linimasa Pengerjaan

V.2 Pengujian dan Evaluasi

Pengujian dilakukan dengan cara membuat berkas PDF menggunakan Microsoft Word secara manual dengan menerapkan berbagai penempatan posisi elemen dan *font* dengan berbagai ukuran. Setelah itu, berkas PDF dikonversi ke dalam format gambar yang digunakan untuk pengujian. Pengujian dilakukan menggunakan pustaka PyMuPDF untuk memperoleh *bounding box* pada dokumen asli dan dokumen keluaran. Selain itu, dilakukan pula pengujian secara manual dengan mendata *font* yang digunakan, seperti jenis dan ukurannya, untuk mengecek kesesuaiannya berdasarkan yang dipakai pada dokumen saat penulisan menggunakan Microsoft Word. Pengujian dilakukan terhadap 100 halaman dengan konten yang berbeda-beda.

V.3 Analisis Risiko dan Mitigasi

Pengembangan sistem rekonstruksi dokumen fisik ke format digital menghadapi berbagai risiko yang dapat memengaruhi keberhasilan implementasi. Risiko-risiko tersebut mencakup aspek teknis seperti keterbatasan akurasi model, ketersediaan sumber daya komputasi, hingga tantangan dalam menangani variasi dokumen yang kompleks. Untuk mengantisipasi hal tersebut, diperlukan identifikasi risiko secara sistematis beserta strategi mitigasi yang tepat. Tabel V.4 menyajikan daftar risiko potensial yang mungkin terjadi selama pengembangan sistem beserta langkah-langkah mitigasi yang akan diterapkan untuk meminimalkan dampak negatif terhadap pencapaian tujuan Tugas Akhir ini.

Tabel V.4 Analisis Risiko dan Strategi Mitigasi

No	Risiko	Strategi Mitigasi
1	Akurasi PP-StructureV3 tidak memenuhi target minimal 75% pada dokumen tertentu	Melakukan *preprocessing* tambahan seperti peningkatan kualitas gambar dan koreksi orientasi. Menyiapkan *fallback mechanism* dengan *pipeline* alternatif. Membatasi jenis dokumen yang diproses sesuai kapabilitas model
2	Hasil OCR yang buruk atau kurang akurat dari bawaan PP-StructureV3	Penambahan pendekatan berbasis VLM pada *pipeline* untuk meningkatkan akurasi ekstraksi teks. Melakukan *post-processing* dengan teknik koreksi ejaan dan validasi konteks
3	Variasi kualitas dokumen *input* yang kurang sesuai seperti adanya *background* sehingga dimensi dokumen terlalu lebar atau terlalu tinggi yang menyebabkan penentuan *bounding box* tidak tepat	Menerapkan deteksi tepi dokumen (*edge detection*) untuk melakukan *cropping* otomatis sehingga diperoleh area dokumen yang presisi. Melakukan normalisasi ukuran dokumen sesuai standar rasio aspek halaman
4	Dokumen dengan elemen kompleks (grafik, diagram) tidak tertangani dengan baik	Fokus pada dokumen dengan struktur standar sesuai batasan masalah. Mendokumentasikan keterbatasan sistem. Merencanakan pengembangan *future work* untuk elemen kompleks

Setiap risiko yang teridentifikasi akan dipantau secara berkelanjutan selama proses pengembangan. Jika terdapat risiko yang terealisasi dan strategi mitigasi awal tidak efektif, akan dilakukan evaluasi ulang untuk menyesuaikan pendekatan atau bahkan kembali ke tahap eksplorasi jika diperlukan. Pendekatan iteratif ini memastikan bahwa sistem yang dikembangkan dapat mencapai tujuan yang telah ditetapkan meskipun menghadapi berbagai tantangan teknis selama implementasi.

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.

Chen, Yufan, Ruiping Liu, Junwei Zheng, Di Wen, Kunyu Peng, Jiaming Zhang, dan Rainer Stiefelhagen. 2025. *Graph-based Document Structure Analysis.* arXiv preprint. arXiv: 2502.02501 [cs.CV]. https://arxiv.org/abs/2502.02501.

Cui, Cheng, Ting Sun, Manhui Lin, Tingquan Gao, Yubo Zhang, Jiaxuan Liu, Xueqing Wang, dkk. 2025. *PaddleOCR 3.0 Technical Report.* arXiv: 2507.05595 [cs.CV]. https://arxiv.org/abs/2507.05595.

Dias, Mariana, dan Carla Teixeira Lopes. 2023. "Optimization of Image Processing Algorithms for Character Recognition in Cultural Typewritten Documents". *Journal on Computing and Cultural Heritage* 16 (4): 1–25. ISSN: 1556-4711. https://doi.org/10.1145/3606705. http://dx.doi.org/10.1145/3606705.

Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, dkk. 2021. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.* arXiv: 2010.11929 [cs.CV]. https://arxiv.org/abs/2010.11929.

Fleischhacker, David, Roman Kern, dan Wolfgang Göderle. 2025. "Enhancing OCR in historical documents with complex layouts through machine learning". *International Journal on Digital Libraries* 26 (3). https://doi.org/10.1007/s00799-025-00413-z.

Ghojogh, Benyamin, dan Ali Ghodsi. 2023. *Recurrent Neural Networks and Long Short-Term Memory Networks: Tutorial and Survey.* arXiv: 2304.11461 [cs.LG]. https://arxiv.org/abs/2304.11461.

Huang, Zhiheng, Wei Xu, dan Kai Yu. 2015. *Bidirectional LSTM-CRF Models for Sequence Tagging.* arXiv: 1508.01991 [cs.CL]. https://arxiv.org/abs/1508.01991.

Islam, Khawar. 2023. *Recent Advances in Vision Transformer: A Survey and Outlook of Recent Work.* arXiv: 2203.01536 [cs.CV]. https://arxiv.org/abs/2203.01536.

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, D. L., S. K. Lee, dan Y. T. Liu. 2025. "Printed document layout analysis and optical character recognition system based on deep learning". *Scientific Reports* 15:23761. https://doi.org/10.1038/s41598-025-07439-y.

Li, Minghao, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, dan Furu Wei. 2022. *TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models.* arXiv: 2109.10282 [cs.CL]. https://arxiv.org/abs/2109.10282.

Ma, Zhichao, Fan Huang, Lu Zhao, Fengjun Guo, Guangtao Zhai, dan Xiongkuo 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.

O'Shea, Keiron, dan Ryan Nash. 2015. *An Introduction to Convolutional Neural Networks.* arXiv: 1511.08458 [cs.NE]. https://arxiv.org/abs/1511.08458.

Ogilvie, Brian. 2016. "Scientific Archives in the Age of Digitization". *Isis* 107 (1): 77–85. https://doi.org/10.1086/686075.

Ouyang, Linke, Yuan Qu, Hongbin Zhou, Jiawei Zhu, Rui Zhang, Qunshu Lin, Bin Wang, dkk. 2025. *OmniDocBench: Benchmarking Diverse PDF Document Parsing with Comprehensive Annotations.* arXiv: 2412.07626 [cs.CV]. https://arxiv.org/abs/2412.07626.

Pfitzmann, Birgit, Christoph Auer, Michele Dolfi, Ahmed S. Nassar, dan Peter Staar. 2022. “DocLayNet: A Large Human-Annotated Dataset for Document-Layout Segmentation”. Dalam *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* 3743–3751. KDD ’22. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3534678.3539043.

Shehzadi, Tahira, Didier Stricker, dan Muhammad Zeshan Afzal. 2024. *A Hybrid Approach for Document Layout Analysis in Document Images.* arXiv preprint. arXiv: 2404.17888 [cs.CV]. https://arxiv.org/abs/2404.17888.

Shi, Baoguang, Xiang Bai, dan Cong Yao. 2015. *An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition.* arXiv: 1507.05717 [cs.CV]. https://arxiv.org/abs/1507.05717.

Sinha, Rasha, dan Rekha B S. 2025. *Digitization of Document and Information Extraction using OCR.* arXiv preprint. arXiv: 2506.11156 [cs.CV]. https://arxiv.org/abs/2506.11156.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, dan Illia Polosukhin. 2023. *Attention Is All You Need.* arXiv: 1706.03762 [cs.CL]. https://arxiv.org/abs/1706.03762.

Wang, Zhangyang, Jianchao Yang, Hailin Jin, Eli Shechtman, Aseem Agarwala, Jonathan Brandt, dan Thomas S. Huang. 2015. *DeepFont: Identify Your Font from An Image.* arXiv: 1507.03196 [cs.CV]. https://arxiv.org/abs/1507.03196.

Wei, Haoran, Yaofeng Sun, dan Yukun Li. 2025. *DeepSeek-OCR: Contexts Optical Compression.* arXiv: 2510.18234 [cs.CV]. https://arxiv.org/abs/2510.18234.

Xiong, Eugene. May 11, 2021. “The Sustainable Impact Of A Paperless Office”. Forbes Technology Council. Accessed November 17, 2025, May 11, 2021. https://www.forbes.com/councils/forbestechcouncil/2021/05/11/the-sustainable-impact-of-a-paperless-office/.

Yousufi, Mahmood Khan. 2023. “Exploring paperless working: A step towards low carbon footprint”. *European Journal of Sustainable Development Research* 7 (4): em0228. https://doi.org/10.29333/ejosdr/13410.

Zabukovšek, Simona Sternad, Sandra Jordan, dan Samo Bobek. 2023. “Managing Document Management Systems’ Life Cycle in Relation to an Organization’s Maturity for Digital Transformation”. *Sustainability* 15 (21): 15212. https://doi.org/10.3390/su152115212.

Zhang, Qintong, Bin Wang, Victor Shea-Jay Huang, Junyuan Zhang, Zhengren Wang, Hao Liang, Conghui He, dan Wentao Zhang. 2025. *Document Parsing Unveiled: Techniques, Challenges, and Prospects for Structured Information Extraction.* arXiv: 2410.21169 [cs.MM]. https://arxiv.org/abs/2410.21169.

Zhao, Zhiyuan, Hengrui Kang, Bin Wang, dan Conghui He. 2024. *DocLayout-YOLO: Enhancing Document Layout Analysis through Diverse Synthetic Data and Global-to-Local Adaptive Perception.* arXiv: 2410.12628 [cs.CV]. https://arxiv.org/abs/2410.12628.