

IV ANALISIS DAN DESAIN

Bab ini menguraikan rancangan solusi teknis yang diusulkan untuk restorasi dokumen terdegradasi dengan pendekatan Generative Adversarial Network yang dimodifikasi secara khusus untuk meningkatkan keterbacaan teks oleh model pengenalan tulisan tangan. Pembahasan dimulai dari analisis kebutuhan sistem (fungsional dan non-fungsional), dilanjutkan dengan rancangan arsitektur solusi yang mencakup desain Generator (U-Net), Recognizer (Transformer-based yang frozen), Diskriminator Dual-Modal (CNN+LSTM), dan fungsi loss multi-komponen (Adversarial + L1 + CTC). Selanjutnya, detail implementasi solusi diuraikan meliputi lingkungan eksperimen, persiapan dataset (ground truth untuk pengenalan tulisan tangan dan sintetis untuk GAN), implementasi arsitektur model dalam TensorFlow/Keras, prosedur training dengan optimasi alternating, dan justifikasi pemilihan hyperparameter melalui studi ablasi.

IV.1 Analisis Kebutuhan Sistem

Berdasarkan rumusan masalah dan tujuan penelitian, sistem restorasi dokumen yang akan dibangun harus memenuhi serangkaian kebutuhan fungsional dan non-fungsional. Kebutuhan ini menjadi acuan dalam perancangan arsitektur dan tolok ukur dalam evaluasi akhir, sejalan dengan tahapan DSRM.

IV.1.1 Kebutuhan Fungsional

Kebutuhan fungsional mendefinisikan fungsi atau kemampuan spesifik yang harus dimiliki oleh sistem GAN-HTR untuk dapat melakukan restorasi dokumen terdegradasi secara efektif. Delapan kebutuhan fungsional yang diidentifikasi mencakup tiga kategori utama:

1. Kemampuan Core ML untuk restorasi visual dan pengenalan teks,
2. Kemampuan Processing untuk menangani batch inference dan dataset sintetis,
3. Kemampuan System Integration untuk input/output yang terstandar dan antarmuka yang fleksibel.

Setiap kebutuhan dirancang untuk memastikan sistem tidak hanya secara teknis mampu melakukan restorasi, tetapi juga praktis untuk digunakan dalam lingkungan produksi dan dapat diintegrasikan dengan sistem lain.

Table 1: Ringkasan Functional Requirements Sistem GAN-HTR

Kode	Nama Requirement	Deskripsi dan Detail Spesifikasi
FR-1	Kemampuan Restorasi Gambar	Menerima input gambar dokumen terdegradasi dan menghasilkan output gambar versi restorasi yang lebih bersih secara visual.
FR-2	Penanganan Berbagai Jenis Degradasi	Menangani setidaknya 4 jenis degradasi utama: tembusan tinta (*bleed-through*), pemudaran (*fading*), noda (*stains*), dan efek buram (*blur*).
FR-3	Integrasi dengan Proses HTR	Gambar hasil restorasi dapat diproses oleh model pengenalan tulisan tangan untuk menghasilkan transkripsi teks.
FR-4	Kemampuan Inference Batch Processing	Memproses dokumen terdegradasi ukuran penuh secara batch dengan output: (1) Gambar dokumen restorasi penuh, (2) File transkripsi teks (.txt), (3) File metadata (.json) dengan waktu proses dan confidence scores.
FR-5	Dukungan untuk Dataset Sintetis	Menyertakan pipeline untuk pembuatan dataset sintetis terdegradasi dari gambar bersih sebagai mekanisme training utama.
FR-6	Output Prediksi Terdefinisi	Menghasilkan output dengan format konsisten: (1) Visual: PNG/TIFF 300 DPI grayscale, (2) Text: UTF-8 dengan confidence score 0.0-1.0 per baris, (3) Quality: PSNR/SSIM, (4) Error: error code dan message.
FR-7	Penanganan Input Data Fleksibel	Menangani berbagai kondisi input: (1) Format: JPG, PNG, TIFF, PDF multi-page, (2) Validasi: resolusi minimum 128x1024px, (3) Recovery: skip corrupted files, (4) Preprocessing: konversi grayscale dan noise reduction.
FR-8	Kemampuan Integrasi Sistem	Menyediakan antarmuka integrasi: (1) CLI dengan parameter terdefinisi, (2) Python API untuk import aplikasi, (3) Configuration management YAML/JSON, (4) Structured logging interface.

IV.1.2 Kebutuhan Non-Fungsional

Kebutuhan non-fungsional mendefinisikan kriteria kualitas atau batasan performa yang harus dipenuhi oleh sistem GAN-HTR. Empat kebutuhan non-fungsional yang ditetapkan mengukur aspek kualitatif dan kuantitatif dari performa sistem, meliputi:

1. Kualitas Visual untuk mengukur kesamaan dengan ground truth menggunakan PSNR dan SSIM,
2. Keterbacaan Teks sebagai metrik utama keberhasilan sistem melalui penurunan CER,
3. Efisiensi Waktu untuk memastikan kepraktisan penggunaan,
4. Reprodutifitas untuk menjamin konsistensi dan keandalan hasil eksperimen.

Target performa yang ditetapkan berdasarkan studi literatur dan kebutuhan praktis implementasi di lapangan.

Table 2: Ringkasan Non-Functional Requirements Sistem GAN-HTR

Kode	Nama Requirement	Deskripsi dan Target Performa
NFR-1	Kualitas Visual Restorasi	Kualitas visual gambar hasil restorasi vs ground truth harus mencapai: (1) PSNR rata-rata ≥ 35 dB, (2) SSIM rata-rata ≥ 0.90 pada dataset validasi.
NFR-2	Peningkatan Keterbacaan Teks	Sistem harus meningkatkan akurasi transkripsi HTR secara signifikan: CER menunjukkan penurunan minimal 25% dibandingkan CER pada gambar terdegradasi asli.
NFR-3	Performa Inferensi Waktu	Waktu processing untuk satu gambar dokumen tunggal pada NVIDIA RTX A4000 tidak boleh melebihi 15 detik untuk memastikan kepraktisan penggunaan.
NFR-4	Reproduktifitas	Proses training dan evaluasi harus dapat direproduksi dengan lingkungan perangkat lunak dan dependensi yang didefinisikan secara jelas (melalui file <code>pyproject.toml</code>).

Berdasarkan Tabel 1 dan Tabel 2, sistem GAN-HTR yang akan dikembangkan harus memenuhi total 12 kebutuhan (8 fungsional dan 4 non-fungsional) yang mencakup aspek visual, tekstual, performa, dan integrasi sistem.

IV.2 Rancangan Arsitektur

Rancangan solusi bertujuan untuk membangun sebuah model generatif yang tidak hanya mampu membersihkan gambar dari kerusakan visual, tetapi juga secara eksplisit dioptimalkan untuk menghasilkan gambar yang keterbacaan teksnya maksimal. Untuk mencapai tujuan ini, arsitektur GAN standar dimodifikasi dengan memperkenalkan dua elemen kunci: (1) Diskriminator dual-modal yang menilai koherensi antara gambar dan teks, dan (2) Fungsi loss gabungan yang mencakup loss dari model pengenalan tulisan tangan.

IV.2.1 Gambaran Umum Arsitektur

Arsitektur yang diusulkan terdiri dari tiga komponen utama yang berinteraksi dalam sebuah kerangka adversarial, seperti yang diilustrasikan pada Gambar 1. Ketiga komponen tersebut adalah:

1. Generator (G): Sebuah model dengan arsitektur U-Net yang bertugas menerima gambar dokumen terdegradasi dan menghasilkan versi restorasi dari gambar tersebut.
2. Recognizer (R): Sebuah model HTR berbasis Transformer yang sudah terlatih dan dibekukan (frozen). Tujuannya bukan untuk dilatih, melainkan untuk "membaca" teks dari gambar hasil restorasi dan memberikan sinyal loss terkait keterbacaan.
3. Diskriminator (D): Sebuah model Diskriminator dual-modal yang menjadi inti dari penelitian ini. Tidak seperti diskriminator pada umumnya yang hanya menerima input gambar, diskriminator ini menerima pasangan (*gambar*, *representasi teks*) untuk menilai apakah sebuah gambar "asli" dan apakah teks di dalamnya koheren.

IV.2.2 Desain Generator

- Jalur Encoder: Berfungsi seperti jaringan konvolusi pada umumnya untuk mengekstraksi fitur dari gambar input. Melalui serangkaian blok konvolusi dan operasi *max-pooling*, resolusi spasial gambar secara bertahap dikurangi (downsampling) sementara kedalaman fitur (jumlah channel) ditingkatkan. Proses ini memungkinkan model untuk menangkap informasi kontekstual dari gambar.
- Jalur Decoder: Bertugas untuk merekonstruksi gambar output dari representasi fitur yang telah dipelajari oleh encoder. Melalui serangkaian blok *upsampling* dan konvolusi, resolusi spasial gambar secara bertahap ditingkatkan hingga kembali ke ukuran aslinya.
- Skip Connections: Ini adalah fitur utama dari arsitektur U-Net. Koneksi ini menghubungkan secara langsung feature map dari lapisan di jalur encoder ke lapisan yang bersesuaian di jalur decoder. Dengan meneruskan informasi dari lapisan awal (yang kaya akan detail spasial) ke lapisan akhir, U-Net mampu

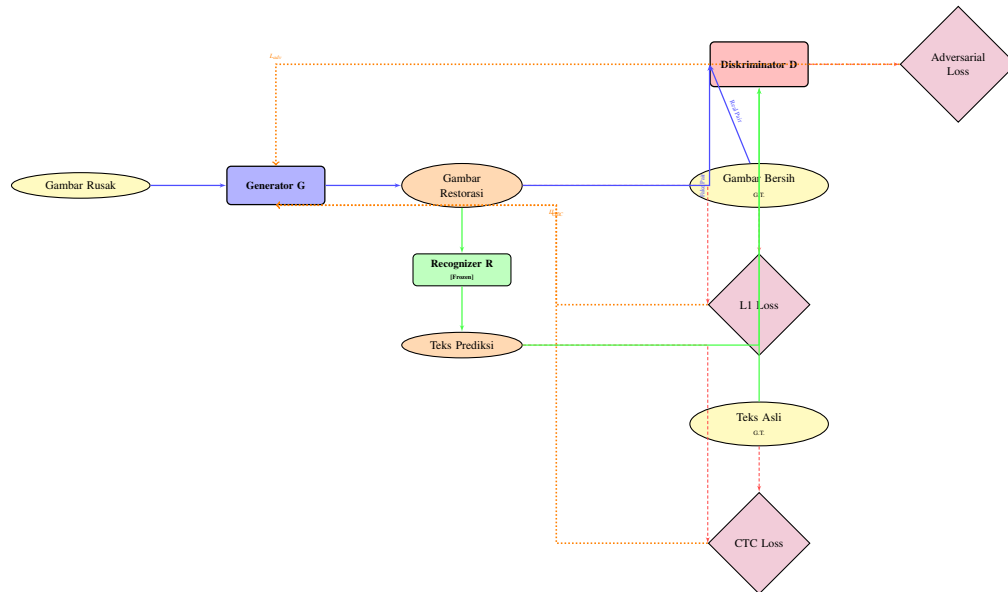


Figure 1: Diagram alur arsitektur GAN yang telah direvisi untuk kejelasan. Diagram ini secara akurat menunjukkan input untuk setiap komponen loss: L1 Loss membandingkan dua gambar, CTC Loss membandingkan dua sekuens teks, dan Adversarial Loss berasal dari Diskriminator.

mengatasi hilangnya informasi detail yang sering terjadi pada arsitektur encoder-decoder standar. Untuk tugas restorasi dokumen, ini sangat penting untuk menjaga ketajaman dan keutuhan gurat-gurat tipis pada karakter tulisan tangan.

Jalur Encoder (Kontraksi)

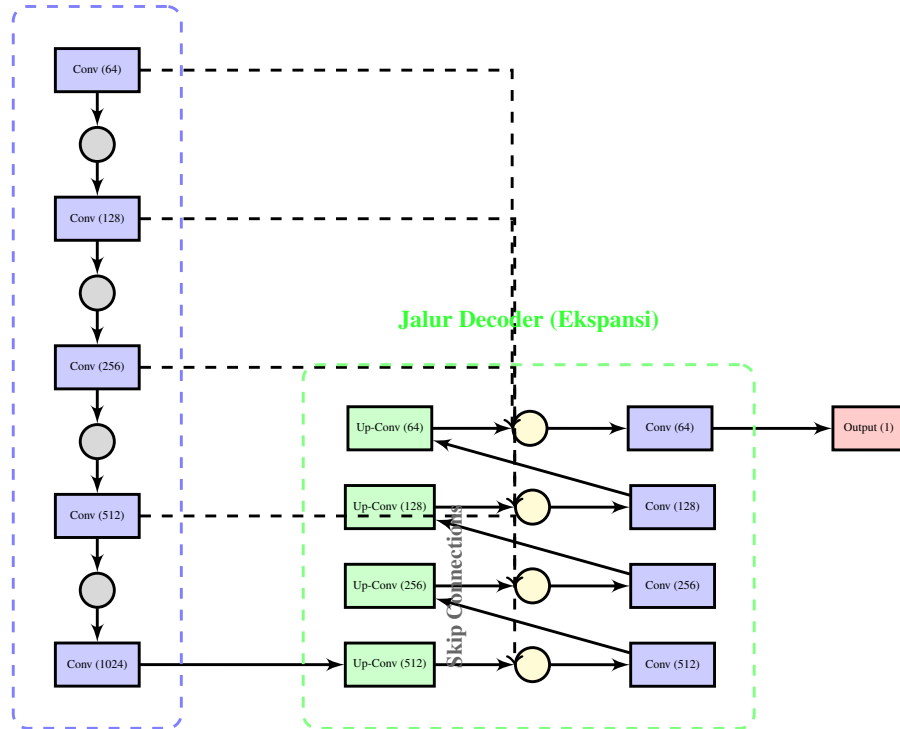


Figure 2: Diagram arsitektur U-Net yang digunakan sebagai Generator.

IV.2.3 Desain Recognizer

Komponen Recognizer (R) memainkan peran khusus dan esensial dalam arsitektur yang diusulkan. Berbeda dengan komponen lain, Recognizer tidak dilatih selama proses training GAN. Sebaliknya, komponen ini merupakan sebuah model Handwritten Text Recognition (HTR) berbasis Transformer yang telah dilatih sebelumnya (*pre-trained*) dan bobotnya dibekukan (*frozen*) atau diatur sebagai *non-trainable*.

Fungsi Utama dan Peran dalam Arsitektur Recognizer berfungsi sebagai evaluator keterbacaan teks yang objektif. Komponen ini menerima gambar hasil restorasi dari Generator dan menghasilkan distribusi probabilitas karakter untuk mengukur keterbacaan hasil restorasi. Output dari Recognizer digunakan untuk dua tujuan:

1. Sebagai input teks untuk Diskriminator, yang menilai koherensi antara gambar dan teks yang dikenali.
2. Sebagai input untuk kalkulasi CTC Loss, yang secara langsung mengukur seberapa dapat dibaca gambar hasil restorasi.

Arsitektur High-Level Recognizer menggunakan arsitektur CNN+Transformer yang dirancang khusus untuk dokumen kuno:

- **CNN Backbone:** Mengekstrak fitur visual hierarkis dari stroke dasar hingga kompleksitas karakter
- **Transformer Encoder:** Menangkap dependensi jarak jauh dan konteks linguistik dengan multi-head attention
- **CTC Output Layer:** Menghasilkan probabilitas karakter untuk setiap timestep

Detail justifikasi pemilihan arsitektur CNN+Transformer untuk kasus paleografi telah dibahas secara komprehensif pada Bab 2.3.4.

IV.2.4 Desain Diskriminator Dual-Modal

Inovasi utama dalam rancangan solusi ini terletak pada arsitektur Diskriminator (D). Berbeda dengan diskriminator pada GAN konvensional yang hanya menilai realisme sebuah gambar (unimodal), diskriminator yang diusulkan bersifat **dual-modal**. Tujuannya tidak hanya untuk membedakan gambar asli dari gambar palsu, tetapi juga untuk menilai **koherensi** antara konten visual sebuah gambar dengan konten semantik dari teks yang diekstraksi dari gambar tersebut.

Diskriminator ini menerima sepasang input: sebuah gambar dan sebuah representasi teks. Tugasnya adalah menghasilkan skor probabilitas tunggal yang menyatakan apakah pasangan tersebut adalah pasangan (*gambar bersih, teks asli*) yang koheren, atau pasangan (*gambar restorasi, teks prediksi*) yang kemungkinan besar tidak koheren.

Arsitektur ini terdiri dari dua jalur pemrosesan paralel yang kemudian digabungkan:

- **Jalur Input Gambar:** Serangkaian lapisan Conv2D dengan LeakyReLU dan BatchNormalization untuk mengekstrak fitur visual
- **Jalur Input Teks:** Layer Embedding dan LSTM untuk memproses representasi teks dan menangkap konteks
- **Fusi dan Klasifikasi:** Konkatenasi fitur visual dan tekstual, diikuti Dense layers untuk klasifikasi akhir

IV.2.5 Desain Fungsi Loss

Optimalisasi arsitektur GAN yang diusulkan mengandalkan dua fungsi loss yang bekerja secara simultan. Desain fungsi loss untuk Generator menjadi salah satu kontribusi utama karena secara eksplisit memasukkan metrik keterbacaan teks.

Fungsi Loss Diskriminator Fungsi loss untuk Diskriminator (L_D) menggunakan Binary Cross-Entropy (BCE) standar untuk membedakan pasangan data asli dan palsu:

$$L_D = -\mathbb{E}_{x,y}[\log D(x,y)] - \mathbb{E}_{z,y'}[\log(1 - D(G(z),y'))] \quad (1)$$

Fungsi Loss Generator Fungsi loss untuk Generator (L_G) dirancang untuk mencapai dua tujuan: menghasilkan gambar realistis dan gambar yang teksnya dapat dibaca dengan baik. L_G disusun sebagai kombinasi tiga komponen loss:

1. Adversarial Loss (L_{adv}) Mengukur kemampuan Generator dalam "menipu" Diskriminator:

$$L_{adv} = -\mathbb{E}_{z,y'}[\log D(G(z),y')] \quad (2)$$

2. Reconstruction Loss (L_{L1}) L1 Loss untuk kemiripan piksel dengan ground truth:

$$L_{L1} = \mathbb{E}_{x,G(z)}[\|x - G(z)\|_1] \quad (3)$$

3. CTC Readability Loss (L_{CTC}) Inovasi kunci: CTC loss dari frozen HTR model untuk mengoptimalkan keterbacaan teks:

$$L_{CTC} = \mathbb{E}_{z,t}[\text{CTC}(R(G(z)),t)] \quad (4)$$

Fungsi Loss Gabungan Ketiga komponen loss tersebut digabungkan menjadi satu fungsi loss total untuk Generator:

$$L_G = \lambda_{adv}L_{adv} + \lambda_{L1}L_{L1} + \lambda_{CTC}L_{CTC} \quad (5)$$

Bobot λ_{adv} , λ_{L1} , dan λ_{CTC} adalah hyperparameter yang dapat disesuaikan untuk menyeimbangkan kontribusi dari setiap komponen loss terhadap proses training Generator.

V KESIMPULAN DAN REKOMENDASI

Berdasarkan analisis kebutuhan sistem dan rancangan arsitektur yang telah diuraikan, bab ini telah menyajikan desain komprehensif untuk framework GAN-HTR yang mampu melakukan restorasi dokumen terdegradasi dengan orientasi pada keterbacaan teks. Rancangan solusi ini memenuhi semua kebutuhan fungsional dan non-fungsional yang telah diidentifikasi, dengan memberikan perhatian khusus pada keseimbangan antara kualitas visual dan keterbacaan tekstual.

V.1 Kesimpulan Desain

Analisis Kebutuhan Sistem Telah diidentifikasi total 12 kebutuhan sistem yang terdiri dari 8 kebutuhan fungsional dan 4 kebutuhan non-fungsional. Kebutuhan-kebutuhan ini mencakup aspek kemampuan restorasi visual, penanganan berbagai jenis degradasi, integrasi dengan HTR, pemrosesan batch, dukungan dataset sintetis, format output yang terdefinisi, fleksibilitas input, kapabilitas integrasi sistem, serta target performa untuk kualitas visual, keterbacaan teks, efisiensi waktu, dan reproduktifitas.

Rancangan Arsitektur Arsitektur yang diusulkan terdiri dari tiga komponen utama yang berinteraksi dalam kerangka kerja adversarial:

1. **Generator (U-Net):** Menerima gambar dokumen terdegradasi dan menghasilkan versi restorasi dengan mempertahankan detail teks melalui skip connections.
2. **Recognizer (Transformer-based HTR):** Model beku yang berfungsi sebagai evaluator keterbacaan teks yang objektif dan konsisten.
3. **Diskriminator Dual-Modal:** Menggabungkan evaluasi visual dan tekstual untuk memastikan hasil restorasi tidak hanya realistis secara visual tetapi juga koheren secara tekstual.

Fungsi Loss Multi-Komponen Fungsi loss gabungan yang mengintegrasikan adversarial loss, L1 reconstruction loss, dan CTC readability loss memberikan optimasi yang seimbang antara kualitas visual dan keterbacaan teks, yang merupakan inovasi kunci dari penelitian ini.

V.2 Rekomendasi Implementasi

Detail implementasi teknis dari rancangan arsitektur ini telah dipindahkan ke Bab 3 (Metodologi Penelitian) pada bagian 3.3.2 "Detail Implementasi Framework". Implementasi mencakup lingkungan eksperimen, pipeline persiapan dataset, konfigurasi training dan hyperparameter, serta sistem evaluasi komprehensif.

Rekomendasi utama untuk implementasi adalah:

- Menggunakan pure FP32 precision untuk stabilitas numerik
- Menerapkan gradient clipping dengan clipnorm=1.0
- Menggunakan bobot loss yang seimbang ($\lambda_{adv} = 1.0$, $\lambda_{L1} = 100.0$, $\lambda_{CTC} = 1.0$)
- Monitor multiple metrics (PSNR, SSIM, CER, WER) untuk evaluasi holistik

V.3 Kontribusi Desain

Rancangan solusi ini memberikan kontribusi pada:

1. **Inovasi Arsitektural:** Pengembangan Dual Modal Discriminator yang secara eksplisit mengevaluasi koherensi visual-teksual.
2. **Integrasi HTR-GAN:** Pendekatan inovatif untuk mengintegrasikan CTC loss dari frozen HTR model ke dalam GAN training.
3. **Optimasi Ganda:** Fungsi loss multi-komponen yang menyeimbangkan antara realism visual dan keterbacaan teks.
4. **Metodologi End-to-End:** Pipeline lengkap dari persiapan dataset hingga evaluasi yang terdokumentasi dengan baik.

Desain ini memberikan fondasi yang kuat untuk implementasi framework GAN-HTR yang dapat mengatasi tantangan restorasi dokumen historis dengan berorientasi pada peningkatan keterbacaan teks, bukan hanya perbaikan visual semata.