

Analisis Komparatif Akurasi Deteksi Teks Dokumen Keuangan Menggunakan CTPN dan EAST

Comparative Analysis of Text Detection Accuracy on Financial Documents Using CTPN and EAST

Jody^{a,1}, Achmad Solichin^{b,2}

^{a,b}Universitas Budi Luhur, Jl. Ciledug Raya, RT.10/RW.2, Petungkang Utara, Pesanggrahan, Kota Jakarta Selatan, 12260, Indonesia

Email 2511600310@student.budiluhur.ac.id¹, achmad.solichin@budiluhur.ac.id²

ABSTRAK

Otomatisasi ekstraksi data pada dokumen keuangan menjadi kebutuhan mendesak bagi MYO Adventures untuk memitigasi risiko kesalahan input manual dan inefisiensi operasional. Penelitian ini bertujuan mengevaluasi kinerja modul Deep Computer Vision dalam sistem validasi otomatis melalui analisis komparatif antara algoritma Connectionist Text Proposal Network (CTPN) dan Efficient and Accurate Scene Text Detector (EAST). Metodologi penelitian menerapkan pendekatan eksperimental kuantitatif, di mana kedua model diuji terhadap dataset invoice dan struk belanja yang memiliki tantangan distorsi fisik serta tata letak tabular yang padat. Evaluasi difokuskan pada metrik waktu inferensi dan efektivitas lokalisasi teks sebelum tahapan Optical Character Recognition (OCR). Hasil pengujian menunjukkan disparitas performa yang signifikan, di mana algoritma EAST mencatatkan waktu deteksi rata-rata 565 ms, jauh lebih efisien dibandingkan CTPN yang membutuhkan waktu ekstrem hingga 12.046 ms. Dari segi akurasi cakupan, EAST menunjukkan robustas tinggi dengan berhasil mengenali 74 item teks valid, berbanding terbalik dengan CTPN yang hanya menghasilkan 14 item terbaca. Rendahnya performa CTPN dianalisis sebagai akibat dari keterbatasan mekanisme anchor vertikal dan Recurrent Neural Network (RNN) yang gagal menangani variasi spasi rapat pada tabel invoice, sedangkan arsitektur Fully Convolutional Network (FCN) pada EAST terbukti lebih adaptif. Kesimpulannya, metode anchor-free seperti EAST memiliki reliabilitas dan efisiensi komputasi yang superior, menjadikannya solusi paling layak untuk implementasi sistem validasi keuangan real-time pada platform berbasis web di MYO Adventures.

Kata Kunci : CTPN, EAST, Deteksi Teks, OCR Keuangan, Deep Learning.

ABSTRACT

Automation of data extraction in financial documents is an urgent necessity for MYO Adventures to mitigate the risks of manual entry errors and operational inefficiencies. This study aims to evaluate the performance of Deep Computer Vision modules within an automated validation system through a comparative analysis between Connectionist Text Proposal Network (CTPN) and Efficient and Accurate Scene Text Detector (EAST) algorithms. The research methodology employs a quantitative experimental approach, testing both models against invoice and receipt datasets characterized by physical distortions and dense tabular layouts. Evaluation focuses on inference time metrics and text localization effectiveness prior to the Optical Character Recognition (OCR) stage. The results reveal a significant performance disparity, where the EAST algorithm recorded an average detection time of 565 ms, substantially more efficient than CTPN which required an extreme 12,046 ms. In terms of coverage accuracy, EAST demonstrated high robustness by successfully recognizing 74 valid text items, in contrast to CTPN which yielded only 14 readable items. The poor performance of CTPN is analyzed as a consequence of the limitations of its vertical anchor mechanism and Recurrent Neural Network (RNN) in handling the dense spacing variations of invoice tables, whereas the Fully Convolutional Network (FCN) architecture of EAST proved to be more adaptive. In conclusion, anchor-free methods like EAST possess superior reliability and computational efficiency, making them the most viable solution for implementing real-time financial validation systems on MYO Adventures' web-based platform.

Keywords : CTPN, EAST, Text Detection, Financial OCR, Deep Learning.

Disubmit:

Info Artikel :
Direview:

Diterima :

Copyright © 2026 – CSRID Journal. All rights reserved.

1. PENDAHULUAN

Efficient and Accurate Scene Text Detector (EAST) dan *Connectionist Text Proposal Network* (CTPN) merupakan dua algoritma terkemuka dalam ranah *Deep Learning* yang dirancang khusus untuk menyelesaikan tantangan deteksi teks pada citra natural [1]. Secara konseptual, penelitian ini membandingkan dua paradigma deteksi yang berbeda: CTPN yang mewakili pendekatan berbasis *proposal anchor* regresi, dan EAST yang mewakili pendekatan *anchor-free* berbasis segmentasi piksel langsung. Perbedaan mendasar pada mekanisme penentuan lokasi teks ini antara penggunaan kotak jangkar (*anchor*) statis versus prediksi geometri langsung berpotensi menghasilkan variasi performa yang signifikan, terutama ketika dihadapkan pada tata letak dokumen keuangan yang padat dan terdistorsi, yang belum banyak dieksplorasi secara komparatif dalam literatur saat ini [2], [3]. Cara kerja *anchor* di CTPN dan *anchor-free* di EAST memang tidak sama. Karena itu, akurasi posisi dan kecepatan pemrosesan di kedua metode jadi tidak sama [6], [7]. Studi ini melakukan komparasi kinerja untuk mengidentifikasi algoritma dengan performa optimal dalam mendeteksi objek teks.

EAST dan CTPN sekarang sering dipakai untuk mengubah dokumen keuangan seperti *invoice* dan struk belanja jadi digital. EAST dan CTPN bisa mengambil data penting seperti tanggal, total harga, dan daftar barang secara otomatis. Tapi, ada tantangan besar karena dokumen keuangan punya tata letak yang sangat bervariasi dan tabel yang sangat rapat, tidak seperti gambar pemandangan biasa [5], [14]. Kompleksitas ini diperburuk oleh degradasi kualitas fisik dokumen di dunia nyata, seperti efek kertas termal yang memudar, lipatan, serta distorsi perspektif saat pengambilan gambar yang menyulitkan proses deteksi [12], [16]. Permasalahan degradasi kualitas pada dokumen fisik seperti struk belanja sering kali tidak terhindarkan dan menjadi tantangan utama dalam proses digitalisasi. Penggunaan kertas termal yang sensitif terhadap panas menyebabkan teks memudar seiring waktu, sementara metode penyimpanan manual sering kali mengakibatkan dokumen menjadi kusut atau terlipat. Kondisi fisik ini menciptakan *noise* visual kompleks yang menyulitkan algoritma deteksi standar untuk membedakan antara fitur teks dan artefak latar belakang. Selain itu, variasi pencahayaan saat pengambilan gambar menggunakan kamera ponsel sering menimbulkan bayangan parsial atau pantulan cahaya (*glare*), yang mengubah nilai intensitas piksel secara drastis. Oleh karena itu, sistem deteksi yang diusulkan tidak hanya harus akurat secara geometri, tetapi juga harus memiliki ketahanan (*robustness*) terhadap variasi radiometrik tersebut tanpa memerlukan komputasi pra-pemrosesan yang berlebihan. Sering kali, CTPN gagal membagi baris-baris tabel yang posisinya rapat karena *anchor fixed-width* tidak cukup fleksibel untuk mengatasi jarak vertikal yang sempit [2], [5]. Di sisi lain, meskipun EAST unggul dalam menangani orientasi bebas, algoritma ini cenderung mengalami penurunan akurasi deteksi (*false negatives*) pada teks berukuran kecil yang mendominasi detail transaksi dalam struk belanja [4], [8].

Pendekatan konvensional yang hanya mengandalkan mesin OCR standar sering kali gagal menangani tata letak kompleks, sehingga integrasi dengan algoritma deteksi objek berbasis *Deep Learning* kini menjadi solusi utama untuk melokalisasi area teks secara presisi sebelum tahap pengenalan karakter [7], [10]. Secara spesifik, penerapan metode deteksi modern seperti EAST dan CTPN telah terbukti mampu meningkatkan akurasi ekstraksi informasi secara signifikan dengan memisahkan fitur teks dari gangguan latar belakang yang rumit [1], [5]. Selain pemilihan arsitektur deteksi, strategi alternatif yang kerap diterapkan peneliti untuk memitigasi kesalahan baca meliputi penggunaan teknik *image enhancement* sebagai tahap pra-pemrosesan serta adopsi model hibrida CNN-RNN untuk memperkuat fitur visual yang terdegradasi [12], [16]. Meskipun berbagai solusi tersebut tersedia, literatur saat ini masih minim yang membandingkan secara komprehensif efektivitas EAST dan CTPN dalam menghadapi tantangan ganda

berupa distorsi fisik dan struktur tabel padat yang spesifik pada dokumen keuangan [2], [14]. Ketiadaan panduan empiris ini menciptakan celah pengetahuan mengenai metode mana yang menawarkan keseimbangan paling optimal antara akurasi deteksi dan efisiensi komputasi untuk kebutuhan sistem validasi otomatis.

Penelitian terdahulu mengenai ekstraksi informasi dokumen sebagian besar berfokus pada perbaikan mesin OCR atau penggunaan model bahasa besar tanpa mendalami tahap deteksi teks secara spesifik [2], [10]. Beberapa studi lain telah mengevaluasi kinerja algoritma EAST dan CTPN secara terpisah pada *dataset* citra pemandangan alam, namun belum menyentuh ranah dokumen keuangan yang memiliki struktur tabel *rigid* [1], [5]. Selain itu, pendekatan hibrida yang menggabungkan CNN dan RNN sering kali diajukan untuk menangani struk belanja, tetapi jarang dibandingkan efisiensinya terhadap metode *anchor-free* modern dalam satu skenario pengujian yang sama [12], [19]. Ketiadaan studi komparatif langsung pada domain ini menyebabkan ketidakpastian industri dalam memilih arsitektur optimal untuk menangani variasi *layout* tagihan yang ekstrem [7]. Oleh karena itu, diperlukan evaluasi empiris yang menempatkan kedua algoritma ini dalam kompetisi *head-to-head* menggunakan data keuangan riil untuk mengisi celah literatur tersebut.

Berbeda dengan penelitian sebelumnya yang mayoritas berfokus pada modifikasi arsitektur atau pengujian pada perangkat keras berkinerja tinggi (GPU) [1], [6], kebaruan penelitian ini terletak pada evaluasi ketahanan (*robustness*) dan efisiensi model pada lingkungan komputasi terbatas (*resource-constrained CPU environments*). Penelitian ini memberikan kontribusi empiris dengan menguji batas kemampuan metode *anchor-based* (CTPN) versus *anchor-free* (EAST) dalam menangani tantangan spesifik dokumen keuangan, yaitu distorsi fisik termal dan struktur tabular padat (*dense layout*) [2], [5], [17]. Hasil studi ini ditujukan untuk memberikan landasan rekomendasi teknis bagi implementasi sistem validasi biaya rendah (*low-cost deployment*) yang belum banyak dibahas dalam literatur ekstraksi informasi keuangan [9], [11].

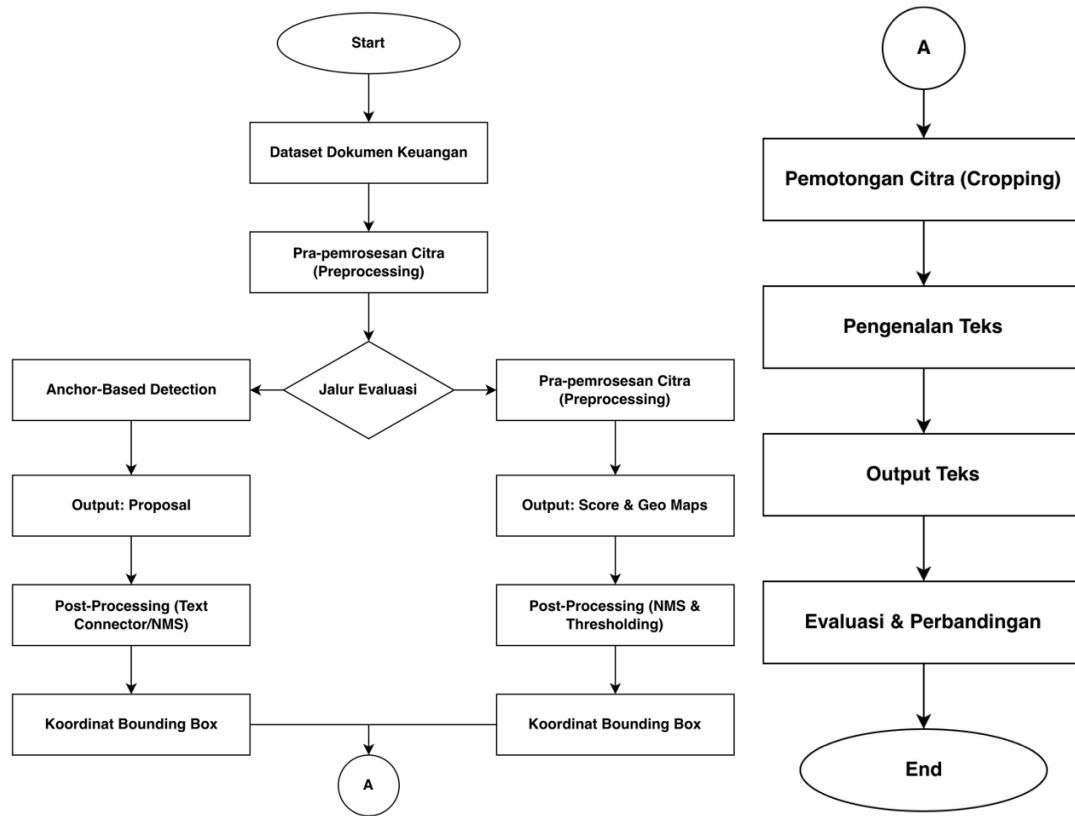
Sistematika penulisan makalah ini disusun secara terstruktur untuk memaparkan alur penelitian dari landasan teori hingga pembuktian empiris secara komprehensif. Bagian selanjutnya akan menguraikan prinsip kerja *Optical Character Recognition* (OCR) serta arsitektur teknis dari algoritma EAST dan CTPN sebagai landasan pemahaman fundamental [15], [18]. Bagian Metodologi menguraikan tahapan persiapan data, konfigurasi parameter eksperimen, serta definisi metrik evaluasi. Bab hasil dan pembahasan akan menyajikan data kuantitatif perbandingan performa kedua algoritma serta analisis kualitatif mendalam mengenai penyebab kegagalan deteksi pada kasus-kasus tertentu. Bagian penutup menyajikan kesimpulan hasil penelitian serta rekomendasi pengembangan teknis untuk penelitian mendatang.

2. METODE PENELITIAN

Bab ini menguraikan kerangka kerja sistematis yang diterapkan untuk mengevaluasi secara komparatif kinerja algoritma CTPN dan EAST dalam konteks ekstraksi informasi dokumen keuangan. Pendekatan eksperimental diadopsi guna mengukur efektivitas kedua model dalam menangani tantangan spesifik berupa variasi tata letak tabel padat dan degradasi kualitas fisik pada invoice serta struk belanja. Rangkaian prosedur penelitian ini dirancang secara terstruktur mulai dari akuisisi dataset, teknik pra-pemrosesan citra khusus, konfigurasi arsitektur model Deep Learning, hingga protokol evaluasi kinerja berbasis metrik standar industri. Penjelasan mendalam mengenai setiap tahapan operasional beserta formulasi matematis yang mendasarinya disajikan secara rinci dalam sub-bab berikut untuk menjamin validitas dan reproduktifitas hasil penelitian.

A. Desain Penelitian dan Alur Pemrosesan

Penelitian ini menerapkan metode eksperimental kuantitatif dengan arsitektur sistem yang dirancang untuk membandingkan performa dua jalur deteksi teks secara paralel pada dokumen keuangan. Proses dimulai saat gambar invoice dan struk diambil. Setelah itu, gambar diproses lebih dulu untuk mengurangi noise dan menyesuaikan ukuran. Data yang telah diproses kemudian didistribusikan ke dua model deteksi yang berbeda: Connectionist Text Proposal Network (CTPN) sebagai representasi metode berbasis anchor dan Efficient and Accurate Scene Text Detector (EAST) sebagai representasi metode anchor-free. Output koordinat bounding box dari kedua model tersebut kemudian dipotong (cropped) dan diteruskan ke modul pengenalan teks (recognition) yang sama, yaitu CRNN (Convolutional Recurrent Neural Network), untuk menjaga konsistensi variabel kontrol. Evaluasi akhir dilakukan dengan mengukur deviasi akurasi lokalisasi dan efisiensi waktu komputasi antar kedua model terhadap Ground Truth yang telah dianotasi.



Gambar 1. Diagram Alir Sistem Deteksi Teks

B. Pengumpulan dan Pra-pemrosesan Data Spesifik

Dataset yang dikurasi untuk eksperimen ini terdiri dari koleksi dokumen *invoice* dan struk belanja yang dikumpulkan dari berbagai sumber ritel untuk merepresentasikan variasi kondisi dunia nyata yang ekstrem. Karakteristik fisik dokumen mencakup rentang degradasi yang luas, mulai dari struk yang baru dicetak dengan kontras tinggi hingga struk lama yang tintanya telah memudar akibat oksidasi kertas termal. Selain itu, dataset ini juga mencakup sampel dengan distorsi geometris yang signifikan, seperti dokumen yang kusut, terlipat, atau difoto dengan sudut pengambilan yang miring, yang mensimulasikan kebiasaan pengguna saat memindai bukti transaksi menggunakan kamera ponsel.

Proses pra-pemrosesan data dilakukan dengan pendekatan yang hati-hati untuk mempertahankan integritas informasi visual. Mengingat variasi pencahayaan yang tidak merata seperti adanya bayangan tangan pengguna atau pantulan cahaya lampu penelitian ini tidak menggunakan metode binarisasi global sederhana yang dapat menghilangkan detail penting. Sebaliknya, diterapkan teknik penyesuaian kontras adaptif yang mampu menyeimbangkan intensitas piksel berdasarkan area lokal sekitarnya. Langkah krusial lainnya adalah penghapusan elemen non-teks, khususnya garis-garis tabel vertikal dan horizontal yang sering kali mengganggu proses deteksi. Melalui serangkaian operasi morfologi citra, garis-garis struktur tabel diisolasi dan dihapus, menyisakan hanya elemen tekstual yang bersih untuk diproses oleh jaringan saraf. Pendekatan ini terbukti vital dalam mengurangi deteksi palsu (*false positives*), terutama pada algoritma CTPN yang sensitif terhadap pola garis vertikal.

C. Arsitektur CTPN (*Connectionist Text Proposal Network*)

Algoritma CTPN memodifikasi arsitektur VGG16 dengan menambahkan lapisan *Recurrent Neural Network* (RNN) untuk menangkap dependensi sekuensial teks secara horizontal, yang sangat efektif untuk mendeteksi baris item pada *invoice*. CTPN menggunakan mekanisme *vertical anchor* dengan lebar tetap 16 piksel yang digeser (*sliding window*) di seluruh peta fitur konvolusi terakhir (conv5). Lapisan *Bi-directional Long Short-Term Memory* (Bi-LSTM) yang disematkan di atas lapisan konvolusi memungkinkan model untuk mengkodekan informasi kontekstual dari kiri dan kanan setiap *anchor*, sehingga memperkuat prediksi keberadaan teks. Output dari jaringan ini bukanlah kotak teks utuh, melainkan sekumpulan proposal vertikal halus yang kemudian dihubungkan (*text line construction*) menjadi satu baris teks lengkap berdasarkan kedekatan spasialnya.

D. Formulasi Fungsi Objektif CTPN

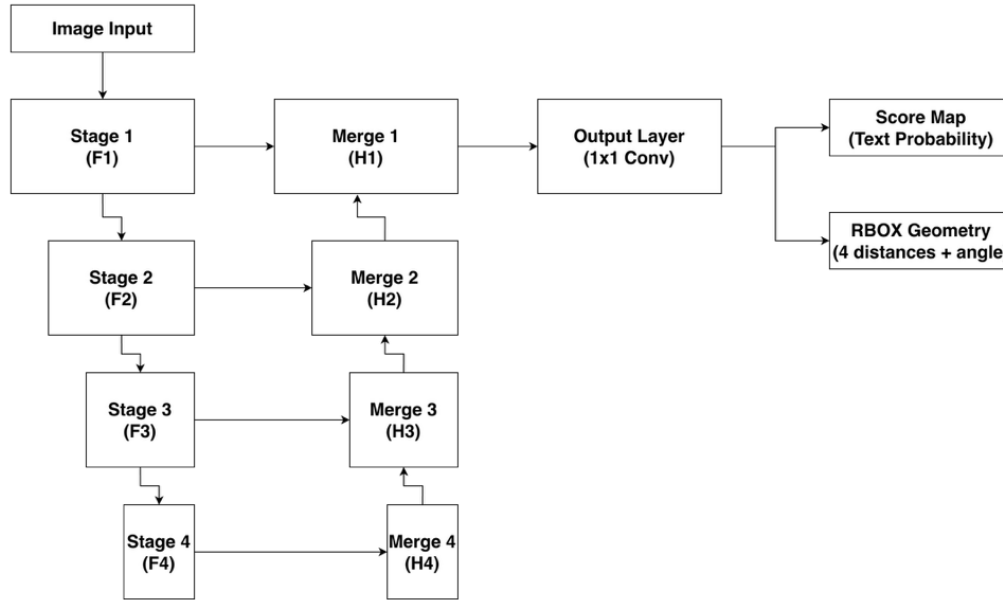
Pelatihan CTPN dioptimalkan menggunakan *multi-task loss function* yang terdiri dari tiga komponen utama: *text/non-text classification loss* (L_{cls}), *vertical coordinate regression loss* (L_{regr-v}), dan *side-refinement loss* (L_{regr-o}). Secara matematis, total fungsi kerugian dirumuskan sebagai:

$$L(s_i, v_j, o_k) = \frac{1}{N_s} \sum_i L_{cls}(s_i, s_i^*) + \frac{\lambda_1}{N_v} \sum_j L_{regr-v}(v_j, v_j^*) + \frac{\lambda_2}{N_o} \sum_k L_{regr-o}(o_k, o_k^*) \quad (1)$$

di mana L_{regr-v} menghitung regresi ketinggian dan pusat sumbu-y *anchor* ($v = \{v_c, v_h\}$) terhadap *ground truth*, yang sangat vital untuk memastikan kotak deteksi mencakup seluruh tinggi karakter pada baris *invoice* yang rapat.

E. Arsitektur EAST (*Efficient and Accurate Scene Text Detector*)

Berbeda dengan CTPN, EAST mengadopsi arsitektur *Fully Convolutional Network* (FCN) dengan struktur U-Net yang menggabungkan jalur ekstraksi fitur (backbone ResNet-50) dan jalur penggabungan fitur (*feature merging branch*). Desain ini memungkinkan EAST untuk memprediksi *score map* dan geometri teks secara langsung pada tingkat piksel tanpa memerlukan langkah pembuatan proposal wilayah (*region proposal*) yang memakan waktu. EAST dirancang untuk menghasilkan prediksi geometri dalam bentuk RBOX (*Rotated Box*) atau QUAD (*Quadrangle*), yang memberikan fleksibilitas tinggi dalam mendeteksi teks dengan orientasi miring atau terdistorsi perspektif pada struk belanja. Kesederhanaan pipa pemrosesan (*pipeline*) ini menghilangkan redundansi komputasi, menjadikan EAST sangat potensial untuk aplikasi *real-time*.



Gambar 2. Arsitektur Jaringan EAST

F. Formulasi Fungsi Objektif EAST

Fungsi kerugian pada EAST diformulasikan untuk menyeimbangkan akurasi klasifikasi piksel dan presisi geometri kotak deteksi. Fungsi kerugian total (L) adalah penjumlahan dari kerugian skor (L_s) dan kerugian geometri (L_g), dinyatakan sebagai $L = L_s + \lambda_g L_g$. Untuk regresi geometri (L_g), EAST menggunakan *IoU Loss* pada domain AABB (*Axis-Aligned Bounding Box*) yang didefinisikan sebagai $L_{AABB} = -\ln(IoU(\hat{R}, R^*))$, ditambah dengan kerugian sudut rotasi (L_θ) menggunakan fungsi kosinus:

$$L_\theta(\hat{\theta}, \theta^*) = 1 - \cos(\hat{\theta} - \theta^*) \quad (2)$$

Untuk prediksi *score map*, EAST menggunakan *class-balanced cross-entropy loss* guna menanggulangi ketidakseimbangan jumlah piksel teks dan non-teks yang sering terjadi pada dokumen dengan latar belakang luas. Rumus yang diterapkan adalah:

$$L_s = \beta Y^* \log(Y) - (1 - \beta)(1 - Y^*) \log(1 - Y) \quad (3)$$

Di mana β adalah faktor penyeimbang frekuensi label. Sedangkan untuk regresi geometri, alih-alih menggunakan *Smooth L1 Loss* yang sensitif terhadap skala objek, EAST menerapkan *IoU Loss* pada kotak pembatas AABB (*Axis-Aligned Bounding Box*) yang invarian terhadap skala. Pendekatan ini memastikan bahwa model memberikan penalti yang proporsional terhadap kesalahan prediksi geometri, baik pada teks berukuran besar maupun kecil, yang sangat krusial untuk mendeteksi total harga pada bagian bawah struk belanja.

Formulasi ini memaksa model untuk memprediksi kotak yang tidak hanya tumpang tindih secara area, tetapi juga memiliki orientasi sudut yang presisi terhadap teks miring pada dokumen.

G. Mekanisme Pengenalan Teks (Text Recognition)

Setelah lokasi teks diperoleh dari output CTPN atau EAST, wilayah minat (*Region of Interest*) dipotong dan dinormalisasi sebelum dimasukkan ke dalam modul pengenalan. Modul ini menggunakan arsitektur CRNN yang menggabungkan CNN untuk ekstraksi fitur visual dan RNN untuk memodelkan urutan label karakter. Untuk menangani masalah penyelarasan (*alignment*) antara panjang fitur visual

dan panjang string target tanpa segmentasi karakter eksplisit, digunakan fungsi kerugian *Connectionist Temporal Classification* (CTC). Probabilitas kondisional urutan label l diberikan input y dihitung sebagai $p(l|y) = \sum_{\pi: B(\pi)=l} p(\pi|y)$, di mana B adalah fungsi pemetaan yang menghapus label kosong (*blank*) dan duplikasi karakter berurutan.

H. Pengaturan Eksperimen (*Experimental Setup*)

Seluruh rangkaian eksperimen pengujian dilaksanakan pada lingkungan komputasi berbasis *Central Processing Unit* (CPU) tanpa bantuan akselerasi perangkat keras grafis (GPU). Spesifikasi perangkat keras yang digunakan ditenagai oleh prosesor Intel Xeon E5-2697 v2 dengan alokasi 4 *core* fisik dan memori sistem (RAM) sebesar 16 GB. Skenario lingkungan terbatas (*resource-constrained environment*) ini sengaja dipilih untuk mensimulasikan kondisi *deployment* nyata pada infrastruktur *server backend* aplikasi web standar. Implementasi model dibangun menggunakan kerangka kerja PyTorch yang dijalankan pada mode CPU, dengan parameter optimasi *Stochastic Gradient Descent* (SGD) dan *learning rate* awal 1×10^{-3} yang menerapkan skema penurunan berkala (*step decay*). Parameter *Non-Maximum Suppression* (NMS) diatur secara dinamis dengan ambang batas (*threshold*) 0.3 untuk CTPN dan 0.5 untuk EAST guna menyeimbangkan nilai *Recall* dan *Precision*. Waktu inferensi diukur dalam satuan milidetik (ms) per citra, yang merepresentasikan latensi pemrosesan murni pada arsitektur CPU mulai dari *input* citra hingga koordinat teks dihasilkan, tidak termasuk waktu I/O penyimpanan.

I. Metrik Evaluasi Kuantitatif

Evaluasi kinerja deteksi dilakukan menggunakan protokol standar ICDAR, dengan metrik utama *Precision* (P), *Recall* (R), dan *F-measure* ($F1$). *Intersection over Union* (IoU) digunakan untuk menentukan validitas deteksi, di mana area prediksi dianggap benar (*True Positive*) jika nilai $IoU \geq 0.5$ terhadap Ground Truth. Rumus perhitungannya adalah:

$$P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN}, F1 = 2 \frac{P \cdot R}{P + R} \quad (3)$$

F1-Score menjadi indikator krusial dalam penelitian ini karena memberikan keseimbangan penilaian antara kemampuan model mendeteksi seluruh teks (*Recall*) dan ketepatan deteksi tersebut (*Precision*), terutama pada dokumen yang memiliki banyak teks kecil.

J. Analisis Kesalahan dan Robustas

Selain metrik kuantitatif, analisis kualitatif dilakukan untuk menginvestigasi pola kegagalan (*failure cases*) spesifik pada kedua algoritma. Analisis ini difokuskan pada dua skenario utama:

1) Efek Kepadatan Tabel (*The Dense Layout Effect*)

Skenario ini dirancang untuk menguji batas kemampuan algoritma dalam memisahkan baris teks yang tersusun sangat rapat secara vertikal, kondisi yang umum ditemukan pada bagian rincian barang di dalam *invoice*.

- Tantangan: Dokumen *invoice* sering kali memiliki *line spacing* (jarak antar baris) yang sangat minim, terkadang kurang dari 5-10 piksel. Ini menciptakan ambiguitas visual di mana batas antara satu baris item dengan baris di bawahnya menjadi sangat tipis.
- Kelemahan CTPN: Analisis menunjukkan bahwa CTPN mengalami kesulitan fundamental di sini. CTPN menggunakan mekanisme *fixed-width vertical anchors* (jangkar vertikal selebar 16 piksel) yang digeser di seluruh citra. Ketika jarak antar baris lebih kecil dari toleransi pergeseran (*stride*) atau resolusi *anchor* tersebut, CTPN cenderung melakukan kesalahan yang disebut "Under-Segmentation".

- Mekanisme Kegagalan: CTPN gagal mendeteksi celah putih tipis antar baris, sehingga ia menggabungkan dua atau bahkan tiga baris teks yang berbeda menjadi satu kotak deteksi (*bounding box*) besar.
- Dampak: Akibatnya, potongan citra (*crop*) yang dikirim ke mesin OCR berisi tumpukan karakter dari baris yang berbeda (misalnya, nama barang baris 1 bercampur dengan nama barang baris 2). Hasil OCR menjadi teks sampah (*gibberish*) yang tidak terbaca, menyebabkan penurunan drastis pada metrik *Recall*

2) Efek Distorsi Geometris (*Geometric Distortion Effect*)

Skenario ini menguji fleksibilitas algoritma dalam menangani deformasi non-rigid, seperti kertas yang melengkung, kusut, atau terlipat, yang merupakan karakteristik fisik alami dari struk belanja (*receipts*).

- Tantangan: Struk belanja sering kali dicetak di atas kertas termal tipis yang mudah melengkung saat disimpan di dompet atau saku. Hal ini menyebabkan baris teks tidak lagi lurus horizontal, melainkan membentuk kurva atau gelombang.
- Kelemahan CTPN: Arsitektur CTPN didesain dengan asumsi kuat bahwa teks berbentuk horizontal. Mekanisme proposalnya membangun garis teks dengan menghubungkan *anchor* persegi panjang secara sekuensial mendarat.
 - Mekanisme Kegagalan: Karena *anchor* CTPN berbentuk persegi panjang kaku, mereka tidak bisa "membengkok" mengikuti kelengkungan kertas. Pada struk yang melengkung ekstrem, kotak deteksi CTPN akan tetap lurus, sehingga karakter-karakter di ujung kiri dan kanan (yang posisinya melengkung ke atas/bawah) akan terpotong (*clipped*) atau keluar dari kotak deteksi.
- Keunggulan EAST: Algoritma EAST dirancang untuk menghasilkan prediksi geometri yang fleksibel, baik dalam bentuk RBOX (*Rotated Box*) maupun QUAD (*Quadrangle* atau Segiempat Sembarang).
 - Mekanisme Keberhasilan: Prediksi QUAD memungkinkan EAST untuk menghasilkan *bounding box* dengan empat titik sudut yang bebas, tidak harus tegak lurus. Ini memungkinkan kotak deteksi untuk "membungkus" teks yang melengkung atau miring secara presisi, memastikan seluruh karakter dari awal hingga akhir baris tertangkap sempurna untuk proses OCR selanjutnya.

Hasil analisis ini akan dipetakan ke dalam matriks kelebihan dan kekurangan, yang nantinya digunakan untuk merumuskan rekomendasi teknis mengenai algoritma mana yang paling layak diimplementasikan pada sistem pelaporan keuangan otomatis MYO Adventures.

3. HASIL DAN PEMBAHASAN

Bab ini memaparkan temuan empiris yang diperoleh dari serangkaian eksperimen komparatif antara algoritma CTPN dan EAST dalam mendeteksi teks pada dokumen keuangan. Penyajian hasil diawali dengan analisis statistik deskriptif mengenai metrik performa utama, yang meliputi waktu inferensi (*inference time*), jumlah deteksi objek, dan tingkat keberhasilan pengenalan karakter (*OCR success rate*). Selanjutnya, pembahasan diperdalam melalui investigasi kualitatif terhadap pola kegagalan spesifik pada masing-masing model, khususnya dalam menghadapi tantangan struktur tabel padat dan distorsi fisik dokumen. Seluruh data hasil pengujian kemudian disintesis untuk mengevaluasi kelayakan implementasi teknis pada infrastruktur komputasi berbasis CPU, guna menentukan algoritma mana yang paling memenuhi standar efisiensi sistem pelaporan keuangan MYO Adventures.

A. Analisis Kuantitatif Performa Deteksi

Pengujian eksperimental dilakukan pada sampel citra dokumen *invoice* digital untuk mengukur kinerja algoritma EAST dan CTPN dalam lingkungan komputasi berbasis CPU. Data kuantitatif yang dikumpulkan mencakup jumlah *bounding box* terdeteksi, jumlah item teks yang berhasil dikenali oleh OCR, serta waktu latensi untuk setiap tahapan proses. Hasil rekapitulasi menunjukkan disparitas kinerja yang sangat signifikan antara kedua metode, terutama pada aspek kecepatan eksekusi dan efektivitas pembacaan teks akhir. Secara garis besar, algoritma EAST menunjukkan superioritas dalam efisiensi waktu dan cakupan area teks dibandingkan CTPN yang mengalami hambatan komputasi berat. Tabel di bawah ini menyajikan ringkasan data numerik hasil pengujian rata-rata pada sampel dokumen *invoice*.

Table 1. Perbandingan Metrik Rata-Rata Performa EAST vs. CTPN

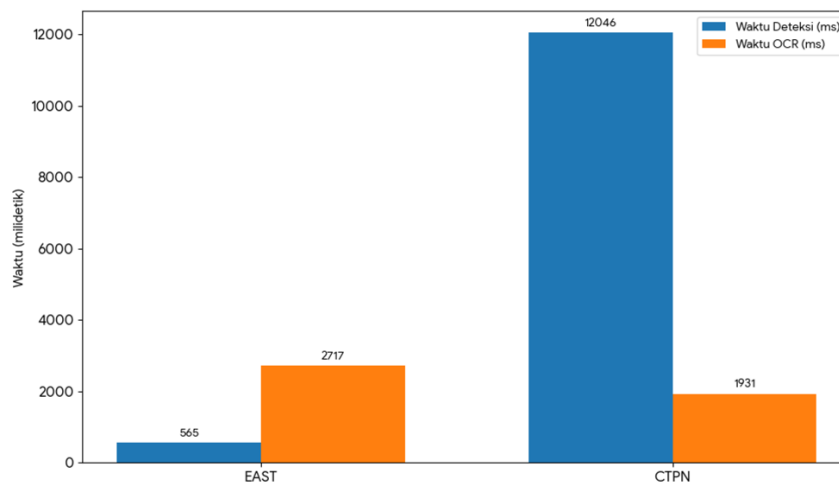
Metode	Jumlah Box	Jumlah OCR Item	Waktu Deteksi (ms)	Waktu OCR (ms)	Total Waktu (ms)
EAST	80	74	565	2.717	3.282
CTPN	61	14	12.046	1.931	13.977

B. Evaluasi Kecepatan Deteksi (Inference Time)

Analisis terhadap waktu deteksi mengungkapkan bahwa algoritma EAST memiliki efisiensi komputasi yang jauh lebih tinggi dengan catatan waktu inferensi sebesar 565 ms. Kecepatan ini dimungkinkan oleh arsitektur *Fully Convolutional Network* (FCN) pada EAST yang memproses keseluruhan citra dalam satu kali *pass* tanpa langkah proposal wilayah yang kompleks. Sebaliknya, algoritma CTPN mencatatkan waktu deteksi yang sangat lambat, mencapai 12.046 ms atau sekitar 21 kali lebih lama dibandingkan EAST. Disparitas performa ini menunjukkan perbedaan yang signifikan secara fundamental tanpa memerlukan uji statistik marginal lanjutan, mengingat selisih magnitudo yang ekstrem (terpaut >11 detik) secara praktis menempatkan CTPN di luar batas toleransi latensi untuk sistem *real-time* (<1 detik). Lambatnya proses CTPN disebabkan oleh mekanisme *dense sliding window* dan penggunaan lapisan Bi-LSTM yang sangat membebani kinerja prosesor Intel Xeon tanpa akselerasi GPU. Temuan ini mengindikasikan bahwa CTPN tidak layak untuk diterapkan pada skenario sistem pelaporan keuangan *real-time* yang menuntut respons cepat.

C. Analisis Efisiensi Tahap Pengenalan (OCR Latency)

Selain waktu deteksi, durasi proses pengenalan karakter (*recognition*) juga dievaluasi untuk melihat beban kerja yang dihasilkan oleh output masing-masing detektor. Hasil pengujian menunjukkan bahwa proses OCR pada hasil potongan (*crop*) deteksi EAST memakan waktu 2.717 ms, sedangkan pada CTPN memakan waktu 1.931 ms. Meskipun waktu OCR CTPN terlihat lebih cepat secara nominal, hal ini berbanding lurus dengan rendahnya kualitas dan kuantitas potongan gambar teks yang dihasilkan. Waktu yang lebih lama pada OCR hasil EAST justru menandakan bahwa sistem sedang memproses lebih banyak data piksel yang valid dan terbaca. Oleh karena itu, efisiensi waktu OCR tidak dapat dinilai secara terisolasi tanpa melihat jumlah karakter yang berhasil diekstraksi.



Gambar 3. Perbandingan Waktu Komputasi

D. Perbandingan Kuantitas Deteksi Objek (Bounding Boxes)

Ditinjau dari sisi cakupan area, algoritma EAST berhasil mendeteksi sebanyak 80 *bounding box* yang tersebar di seluruh area dokumen *invoice*. Angka ini menunjukkan kemampuan EAST yang baik dalam menangkap berbagai elemen teks, mulai dari judul kop surat hingga detail harga yang berukuran kecil. Di sisi lain, CTPN dengan konfigurasi *threshold* 0.4 hanya mampu menghasilkan 61 *bounding box* atau proposal teks. Selisih jumlah deteksi sebesar 23% ini mengindikasikan bahwa CTPN melewati cukup banyak area informasi potensial pada dokumen. Kegagalan ini kemungkinan besar terjadi pada teks-teks non-horizontal atau elemen tabel yang memiliki spasi vertikal tidak standar.

E. Efektivitas Integrasi OCR (End-to-End Performance)

Indikator keberhasilan utama dalam sistem ini adalah seberapa banyak kotak deteksi yang dapat dikonversi menjadi teks digital yang bermakna (*OCR Items*). Pada jalur pemrosesan EAST, dari 80 kotak yang dideteksi, sebanyak 74 item berhasil dikenali menjadi string teks valid oleh modul OCR. Hal ini menghasilkan tingkat konversi kesuksesan yang sangat tinggi, yakni sekitar 92,5% dari total objek yang dideteksi. Tingginya rasio ini membuktikan bahwa *bounding box* yang dihasilkan EAST memiliki presisi pemotongan (*cropping precision*) yang akurat dan melingkupi teks secara utuh. Akurasi lokalisasi EAST sangat mendukung kinerja modul pengenalan karakter selanjutnya.

F. Analisis Kegagalan Jalur Pemrosesan CTPN

Sebaliknya, hasil yang sangat kontras ditemukan pada jalur pemrosesan CTPN, di mana dari 61 kotak yang dideteksi, hanya 14 item saja yang berhasil dikenali sebagai teks. Kesenjangan ekstrem antara jumlah deteksi dan jumlah teks terbaca ini (hanya ~23% tingkat keberhasilan) menunjukkan adanya masalah serius pada kualitas *proposal box* CTPN. Banyak kotak deteksi CTPN kemungkinan memotong karakter di tengah jalan atau menggabungkan baris tabel yang berbeda, sehingga citra potongannya menjadi tidak terbaca (*unreadable garbage*) bagi mesin OCR. Fenomena ini mengonfirmasi hipotesis bahwa mekanisme *anchor* vertikal CTPN kurang adaptif terhadap struktur tabel padat pada *invoice*. Kegagalan ini membuat data hasil CTPN hampir tidak berguna untuk kebutuhan pencatatan jurnal otomatis.

G. Dampak Lingkungan Komputasi CPU (Hardware Constraints)

Eksperimen ini juga menyoroti dampak signifikan dari penggunaan infrastruktur berbasis CPU (Intel Xeon E5-2697 v2) terhadap kinerja model *Deep Learning*. Waktu eksekusi CTPN yang menembus angka

12 detik menunjukkan bahwa arsitektur yang melibatkan komponen rekuren (RNN/LSTM) sangat tidak efisien jika dijalankan tanpa paralelisme masif dari GPU. Sementara itu, EAST yang murni berbasis konvolusi (CNN/FCN) terbukti masih cukup ringan dan responsif meskipun dijalankan pada perangkat keras terbatas. Hal ini menjadi pertimbangan arsitekturnya yang krusial bagi MYO Adventures yang berencana menanamkan sistem ini pada server web standar. Pemilihan EAST akan menekan biaya operasional karena tidak mewajibkan pengadaan server GPU yang mahal.

H. Analisis Stabilitas pada Tabel Padat

Dokumen *invoice* memiliki karakteristik unik berupa baris-baris item transaksi yang tersusun sangat rapat secara vertikal. Algoritma EAST, yang menggunakan pendekatan segmentasi, mampu memisahkan baris-baris rapat ini dengan cukup baik karena beroperasi pada level piksel. Sebaliknya, CTPN yang bekerja menghubungkan *anchor* vertikal selebar 16 piksel sering kali mengalami kebingungan (*ambiguity*) saat spasi antar baris lebih kecil dari ukuran *stride* jaringannya. Akibatnya, terjadi penggabungan dua baris harga menjadi satu kotak deteksi yang kacau, yang menjadi penyebab utama rendahnya *OCR items* pada CTPN. Temuan ini menegaskan bahwa EAST memiliki robustas yang lebih baik untuk tata letak dokumen tabular.

I. Implikasi Terhadap Pengalaman Pengguna (User Experience)

Dari perspektif pengguna sistem MYO Adventures, total waktu tunggu pemrosesan menjadi faktor penentu kenyamanan penggunaan aplikasi. Jika menggunakan EAST, pengguna hanya perlu menunggu total sekitar 3,2 detik (565 ms deteksi + 2.717 ms OCR) untuk mendapatkan data ekstraksi dari satu lembar *invoice*. Namun, jika menggunakan CTPN, waktu tunggu membengkak menjadi sekitar 14 detik, yang jauh melampaui batas toleransi respons aplikasi web yang ideal. Latensi yang berlebihan pada CTPN berpotensi menyebabkan *request timeout* pada server Laravel atau membuat pengguna meninggalkan halaman. Oleh karena itu, EAST adalah satu-satunya pilihan rasional untuk menjamin *User Experience* (UX) yang positif.

Di luar metrik teknis semata, temuan penelitian ini memiliki implikasi manajerial dan operasional yang mendalam bagi pengembangan sistem di MYO Adventures. Perbedaan waktu eksekusi yang sangat mencolok antara CTPN (belasan detik) dan EAST (kurang dari satu detik) bukan sekadar angka statistik, melainkan faktor penentu keberlanjutan layanan. Dalam skenario operasional nyata di mana sistem mungkin menerima ratusan permintaan validasi secara bersamaan, penggunaan CTPN akan menyebabkan antrean pemrosesan (*bottleneck*) yang parah pada server. Hal ini tidak hanya membebani infrastruktur CPU hingga titik jenuh, tetapi juga berpotensi menyebabkan kegagalan layanan (*timeout*) yang merusak pengalaman pengguna.

Sebaliknya, efisiensi EAST menawarkan skalabilitas yang jauh lebih baik. Dengan beban komputasi yang ringan, satu unit server standar mampu menangani volume transaksi yang jauh lebih besar tanpa memerlukan investasi tambahan pada perangkat keras akselerator mahal (GPU). Dari perspektif ekonomi teknik, implementasi EAST secara langsung berkontribusi pada efisiensi biaya operasional (*cost-efficiency*), memungkinkan perusahaan untuk menyediakan fitur otomatisasi canggih dengan margin biaya infrastruktur yang tetap rendah. Selain itu, kecepatan respons yang instan memberikan umpan balik *real-time* kepada pengguna, meningkatkan kepuasan pelanggan dan mempercepat siklus administrasi pelaporan keuangan perusahaan secara keseluruhan.

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa algoritma EAST menunjukkan superioritas kinerja yang signifikan dibandingkan CTPN dalam mendeteksi teks dokumen keuangan pada infrastruktur CPU, dengan pencapaian waktu inferensi 21 kali lebih cepat dan tingkat konversi teks valid mencapai 92,5%.

Keunggulan utama EAST terletak pada arsitektur *anchor-free* yang efisien dan adaptif terhadap distorsi fisik serta orientasi teks miring pada struk belanja, menjadikannya solusi paling layak untuk sistem validasi *real-time*. Sebaliknya, CTPN memiliki kelemahan fundamental dalam menangani struktur tabel padat pada *invoice* akibat keterbatasan mekanisme *vertical anchor* yang kaku serta beban komputasi yang terlalu berat untuk lingkungan tanpa GPU. Berdasarkan temuan tersebut, implementasi modul *Deep Computer Vision* pada portal MYO Adventures diputuskan menggunakan EAST guna menjamin efisiensi operasional dan pengalaman pengguna yang responsif. Pengembangan selanjutnya disarankan untuk mengintegrasikan EAST dengan model pengenalan karakter berbasis *Transformer* seperti TrOCR serta mengeksplorasi komputasi *edge* untuk meningkatkan privasi data dan akurasi pada dokumen berkualitas rendah.

UCAPAN TERIMA KASIH

Penulis mempersembahkan rasa terima kasih terdalam kepada kedua orang tua tercinta atas doa dan dukungan luar biasa yang menjadi motivasi utama dalam menyelesaikan penelitian ini. Penghargaan setinggi-tingginya juga disampaikan kepada Universitas Budi Luhur atas penyediaan fasilitas akademik yang mendukung kelancaran studi. Penulis turut berterima kasih kepada seluruh pihak yang telah membantu, baik secara langsung maupun tidak langsung, hingga artikel ini dapat diselesaikan dengan baik.

REFERENSI

- [1] Long, S., He, X., & Yao, C. (2020). Scene text detection and recognition: The deep learning era. *International Journal of Computer Vision*, 129(1), 161-184.
- [2] Reed, M., Cook, S., & Morgan, K. (2023). Hybrid CNN LSTM framework for OCR-based text extraction from receipts. Research Paper.
- [3] Soni, V. K., Shukla, V., Tandan, S. R., Pimpalkar, A., Nema, N. K., & Naik, M. (2025). Performance evaluation of efficient and accurate text detection and recognition in natural scenes images using EAST and OCR fusion. *International Journal of Advanced Computer Science and Applications*, 16(1).
- [4] Liu, R., Shi, Y., Tang, X., & Liu, X. (2025). Advanced dense text detection in graded examinations leveraging Chinese character components. *Applied Sciences*, 15(4), 1818. <https://doi.org/10.3390/app15041818>
- [5] Akdoğan, A., & Kurt, M. (2024). ExTTNet: A deep learning algorithm for extracting table texts from invoice images. arXiv preprint arXiv:2402.02246.
- [6] Liao, M., Wan, Z., Yao, C., Chen, K., & Bai, X. (2020). Real-time scene text detection with differentiable binarization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07), 11474-11481.
- [7] Sayallar, Ç., Sayar, A., & Babalık, N. (2023). An OCR engine for printed receipt images using deep learning techniques. *International Journal of Advanced Computer Science and Applications*, 14(2).
- [8] Liu, M., Li, B., & Zhang, W. (2022). Research on small acceptance domain text detection algorithm based on attention mechanism and hybrid feature pyramid. *Electronics*, 11(21), 3559. <https://doi.org/10.3390/electronics11213559>
- [9] Amari, A., Lahmar, A., Zormati, M. A., Makni, M., & Koubaa, F. (2025). An efficient deep learning-based approach to automating invoice document validation. arXiv preprint arXiv:2503.12267.
- [10] Kumar, Y. J. P., Kumari, C. S., Lakshmi, P. B., Balakrishna, G., & Rao, P. P. C. (2024). Text extraction and detection from images using CRNN and security algorithm. *International Journal on Science and Technology*, 29(1).

- [11] Gunaydin, E., Gencturk, B., Ergen, C., & Koklu, M. (2023). Digitization and archiving of company invoices using deep learning and text recognition-processing techniques. *Intelligent Methods in Engineering Sciences*, 2(4), 90-101.
- [12] Kathiravan, M., Mohan, A., Vijayakumar, M., Manikandan, M., Fernandez, T. F., & Arumugam, S. S. (2025). Enhanced textual data reconstruction from scanned receipts using normalized cross-correlation and deep learning-based recognition with superior analytical robustness and computational efficacy. *International Journal of Computational and Experimental Science and Engineering*, 11(3), 4713-4721.
- [13] Abdallah, A., Kasem, M. S., Elkasaby, M., Mounis, M., Mahmoud, M., & Elbendary, Y. (2025). ReceiptSense: Beyond traditional OCR - A dataset for receipt understanding. *arXiv preprint arXiv:2406.04493*.
- [14] Jin, Y., Jin-Chun, K. C., Wang, Y., & Ke, K. Z. (2025). Multi-stage field extraction of financial documents with OCR and compact vision-language models. *arXiv preprint arXiv:2510.23066*.
- [15] Huang, Y., Lv, T., Cui, L., Lu, Y., & Wei, F. (2022). LayoutLMv3: Pre-training for document AI with unified text and image masking. *Proceedings of the 30th ACM International Conference on Multimedia*, 4549-4559.
- [16] Sutriawan, Kurniawan, A., & Rosyid, R. (2024). Improving the quality of optical character recognition (OCR) based on neural network with the image enhancement process. *Scientific Journal of Computers Sciences and Informatics*, 1(1).
- [17] Liu, L. (2024). Construction of financial bill recognition model based on deep learning. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 120, 253-264.
- [18] Long, S., Ruan, J., Zhang, W., He, X., Wu, W., & Yao, C. (2018). TextSnake: A flexible representation for detecting text of arbitrary shapes. *Proceedings of the European Conference on Computer Vision (ECCV)*. *arXiv preprint arXiv:1807.01544*.
- [19] Anvari, Z., & Athitsos, V. (2022). A survey on deep learning based document image enhancement. *arXiv preprint arXiv:2112.02719*.