Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer
Vol. 6 No. 2 Juli 2026
p-ISSN: 2827-8135; e-ISSN: 2827-7953; Hal 272-282
DOI: https://doi.org/10.55606/juisik.v6i2.2306
Tersedia: https://journal.sinov.id/index.php/juisik

Pengembangan Model Hibrida CRNN dan *Tesseract* OCR untuk Peningkatan Akurasi Ekstraksi Teks dari Citra Dokumen

Syafrie Abdunnasir Jawad[1], Edy Winarno[2]*, Safuan[3], Muhammad Munsarif [4]
[1-4]Program Studi Informatika, Universitas Muhammadiyah Semarang, Indonesia
Penulis Korespondensi: edywin@unimus.ac.id

Abstract*. Text extraction from document images, especially transaction receipts, remains challenging because receipt images often contain small characters, dense layouts, blur, noise, skew, shadows, and irregular illumination. This study develops a Tri-Expert hybrid OCR model that combines CRNN(H=64), MSF-CRNN(H=64), and Tesseract OCR to improve word-level recognition accuracy on receipt images. The proposed system uses calibrated confidence-based selection with temperature scaling, a CRNN(H=48) rule-based arbiter for conflicting predictions, and a numeric-aware Tesseract fallback for low-confidence numeric tokens. Experiments were conducted on word crops from the CORD dataset and evaluated using Exact Match (EM), Word Error Rate (WER), Character Error Rate (CER), and Character F1-score. On the validation set (N=2,186), the hybrid method achieved EM 91.49%, CER 2.37%, and Char-F1 97.96%. On the test set (N=2,356), it achieved EM 89.52%, CER 3.13%, and Char-F1 97.25%, outperforming both CRNN(H=64) and Tesseract. The results indicate that the proposed hybrid design improves OCR reliability, particularly for numeric and short tokens, while remaining modular and practical for web-based deployment.*

Keywords*: Character Error Rate; Crnn; Optical Character Recognition; Receipt Image; Tesseract*

Abstrak. Ekstraksi teks dari citra dokumen, khususnya struk transaksi, masih menantang karena citra struk sering mengandung karakter kecil, tata letak padat, blur, noise, kemiringan, bayangan, dan pencahayaan tidak merata. Penelitian ini mengembangkan model OCR hibrida berbasis *Tri-Expert* yang mengombinasikan CRNN(H=64), MSF-CRNN(H=64), dan *Tesseract OCR* untuk meningkatkan akurasi pengenalan teks pada level kata. Sistem yang diusulkan menggunakan seleksi berbasis *confidence* yang dikalibrasi dengan *temperature scaling*, *arbiter* berbasis CRNN(H=48) untuk menangani konflik prediksi, serta *numeric-aware fallback* berbasis *Tesseract* untuk token numerik berkepercayaan rendah. Eksperimen dilakukan pada *word-crops* dari dataset CORD dan dievaluasi menggunakan *Exact Match (EM), Word Error Rate (WER), Character Error Rate (CER),* serta *Character F1-score*. Pada data validasi (N=2.186), metode hibrida mencapai EM 91,49%, CER 2,37%, dan Char-F1 97,96%. Pada data uji (N=2.356), metode hibrida mencapai EM 89,52%, CER 3,13%, dan Char-F1 97,25%, mengungguli CRNN(H=64) dan *Tesseract*. Hasil ini menunjukkan bahwa rancangan hibrida mampu meningkatkan reliabilitas OCR, terutama pada token numerik dan kata pendek, serta tetap modular untuk implementasi berbasis web.

Kata kunci: Akurasi OCR; CRNN; Ekstraksi Teks; Struk Transaksi; Tesseract

1. LATAR BELAKANG

Ekstraksi informasi dari struk transaksi masih menjadi masalah praktis pada aplikasi keuangan, *reimbursement*, arsip digital, dan otomasi pencatatan. Berbeda dengan dokumen hasil pemindaian yang relatif stabil, struk biasanya diperoleh melalui kamera ponsel sehingga kualitas citra sangat beragam. Blur, *noise*, kemiringan, bayangan, pencahayaan tidak merata, tekstur kertas, dan kualitas cetak yang pudar membuat karakter kecil pada struk sulit dikenali secara konsisten. Kesalahan satu karakter pada nominal, diskon, pajak, tanggal, atau total dapat mengubah makna informasi transaksi dan menyebabkan proses verifikasi tetap harus dilakukan secara manual.

Naskah Masuk: 19 Maret 2026; Revisi: 24 April 2026; Diterima: 25 Juni 2026; Tersedia: 03 Juni 2026

Sistem OCR untuk struk tidak cukup hanya menghasilkan teks mentah. Pada praktiknya, teks hasil OCR akan dipakai ulang sebagai masukan untuk parsing item, perhitungan total, validasi nominal, atau analitik transaksi. Karena itu, kesalahan pada token numerik dan tanda baca memiliki risiko lebih besar dibanding kesalahan pada token non-kritis. Masalah seperti kesalahan membaca angka 0 sebagai O, angka 1 sebagai I, atau pemisah ribuan/desimal sebagai karakter lain perlu ditangani secara lebih hati-hati.

Pendekatan *deep learning* seperti *Convolutional Recurrent Neural Network* (CRNN) telah menjadi salah satu arsitektur penting dalam pengenalan teks berbasis sekuens karena menggabungkan ekstraksi fitur visual, pemodelan urutan, dan *decoding* berbasis CTC (Shi et al., 2017). Namun, CRNN tunggal tetap memiliki keterbatasan pada teks kecil, buram, simbol, serta karakter yang bentuknya mirip. Di sisi lain, *Tesseract* OCR sebagai engine klasik masih relevan karena ringan, mudah diintegrasikan, dan memiliki dukungan bahasa luas, tetapi performanya sensitif terhadap kualitas citra dan praproses (Smith, 2007; Hegghammer, 2022).

Beberapa model modern berbasis *vision-language* dan *transformer*, seperti ABINet, TrOCR, PARSeq, dan VisionLAN, menunjukkan peningkatan akurasi pada *scene text recognition* dan OCR dokumen (Fang et al., 2021; Li et al., 2021; Bautista & Atienza, 2022; Wang et al., 2021). Kendati demikian, model tersebut umumnya membawa konsekuensi komputasi yang lebih tinggi. Pada skenario struk, satu gambar dapat menghasilkan banyak *word-crops* sehingga biaya inferensi kumulatif menjadi pertimbangan penting. Hal ini membuka ruang bagi pendekatan hibrida yang tetap ringan, tetapi lebih reliabel daripada model tunggal.

Berdasarkan gap tersebut, penelitian ini mengusulkan OCR hibrida berbasis *Tri-Expert* yang memadukan CRNN(H=64), MSF-CRNN(H=64), dan *Tesseract* OCR. Kebaruan utama penelitian terletak pada mekanisme pemilihan prediksi berbasis confidence terkalibrasi, penggunaan CRNN(H=48) sebagai *arbiter* berbasis aturan ketika prediksi antar model berbeda, serta *numeric-aware fallback* berbasis *Tesseract* untuk token numerik tertentu. Dengan desain tersebut, sistem tidak hanya mengejar kenaikan akurasi rata-rata, tetapi juga meningkatkan reliabilitas keputusan pada level kata.

Tujuan penelitian ini adalah merancang dan mengimplementasikan model OCR hibrida untuk pengenalan kata pada citra struk, mengukur peningkatan performanya terhadap *baseline* CRNN(H=64) dan *Tesseract* menggunakan metrik EM, WER, CER, dan Char-F1, serta menilai kontribusi mekanisme fusi terhadap pembacaan token numerik dan kata pendek.

Penelitian menggunakan dataset CORD karena menyediakan anotasi struk dan *word-level ground truth* yang mendukung evaluasi objektif serta replikasi eksperimen (Park et al., 2019).

2. KAJIAN TEORITIS

Optical Character Recognition pada Struk

Optical Character Recognition (OCR) adalah proses mengubah representasi visual teks pada citra menjadi representasi simbolik berupa karakter dan kata. Secara umum, *pipeline* OCR terdiri atas akuisisi citra, praproses, deteksi atau segmentasi area teks, normalisasi, pengenalan teks, dan pascaproses. Pada domain struk, *pipeline* ini lebih sulit karena citra sering berisi teks kecil, jarak antar karakter yang rapat, logo, garis pemisah, *barcode*, dan latar kertas yang tidak bersih.

Praproses citra tetap berperan penting dalam OCR, terutama untuk *engine* klasik. Teknik seperti konversi *grayscale*, peningkatan kontras, peredaman *noise*, binarisasi adaptif, *deskew, resize, padding*, dan normalisasi rasio aspek dapat meningkatkan keterbacaan karakter. Metode *Otsu* dan *Sauvola* banyak digunakan pada binarisasi dokumen, sedangkan CLAHE atau filter median/bilateral dapat membantu ketika pencahayaan dan *noise* tidak seragam (Otsu, 1979; Sauvola & Pietikäinen, 2000).

CRNN, MSF-CRNN, dan Tesseract

CRNN menggabungkan CNN sebagai ekstraktor fitur visual, BiLSTM sebagai pemodel sekuens, dan *Connectionist Temporal Classification* (CTC) sebagai fungsi loss sekaligus mekanisme *decoding* tanpa segmentasi karakter eksplisit. CNN mempelajari pola stroke, tepi, dan bentuk parsial karakter, sedangkan BiLSTM memanfaatkan konteks kiri dan kanan untuk membantu membedakan karakter yang mirip. CTC memungkinkan pelatihan *end-to-end* karena model tidak perlu diberi anotasi batas karakter per posisi (Graves et al., 2006; Hochreiter & Schmidhuber, 1997; Shi et al., 2017).

MSF-CRNN memperkuat CRNN dengan mekanisme *multi-scale* atau *multi-stream feature fusion.* Pada struk, detail stroke tipis dan angka rapat mudah hilang akibat *downsampling* agresif. Dengan memadukan fitur dari beberapa skala atau cabang ekstraksi, MSF-CRNN berpotensi menjaga informasi tepi dan tekstur halus yang penting untuk membedakan karakter kecil (Zou et al., 2023).

Tesseract OCR merupakan *engine open-source* yang telah lama digunakan dan pada versi modern menggunakan LSTM untuk pengenalan teks. Kelebihannya adalah ringan, mudah dipasang, mendukung banyak bahasa, dan cocok untuk dokumen relatif bersih. Namun, performanya dapat menurun pada citra dengan blur, *noise*, distorsi perspektif, atau layout

padat. Karena itu, dalam penelitian ini *Tesseract* diposisikan sebagai *expert* tambahan dan *fallback* numerik, bukan sebagai satu-satunya *recognizer*.

Ensemble, Confidence Calibration, dan Metrik Evaluasi

Ensemble dan mixture of experts memanfaatkan beberapa model dengan karakteristik berbeda untuk mengurangi ketergantungan pada satu prediktor. Prinsip dasarnya adalah *error* antar model tidak selalu terjadi pada sampel yang sama sehingga keputusan gabungan dapat lebih stabil daripada keputusan model tunggal (Dietterich, 2000; Jacobs et al., 1991). Dalam OCR hibrida, penggabungan dapat dilakukan melalui voting, *confidence-based selection*, atau *arbiter* yang memilih prediksi akhir berdasarkan fitur tertentu.

Confidence calibration penting karena *skor confidence* model *deep learning* sering tidak sebanding dengan peluang benar aktual. Model dapat sangat yakin pada prediksi yang salah, sehingga perbandingan *confidence* lintas *expert* menjadi bias. *Temperature scaling* merupakan teknik pascapelatihan yang sederhana dan efektif untuk memperbaiki kalibrasi probabilitas tanpa mengubah arsitektur model (Guo et al., 2017).

Evaluasi OCR pada penelitian ini menggunakan *Exact Match* (EM), *Word Error Rate* (WER), *Character Error Rate* (CER), dan *Character F1-score*. EM mengukur proporsi token yang terbaca sama persis dengan *ground truth*, sedangkan WER pada konteks *word-crop* ekuivalen dengan proporsi kata yang salah. CER dihitung dari jarak edit karakter sehingga sensitif terhadap kesalahan kecil, sedangkan *Char*-F1 menyeimbangkan presisi dan recall pada level karakter (Levenshtein, 1966).

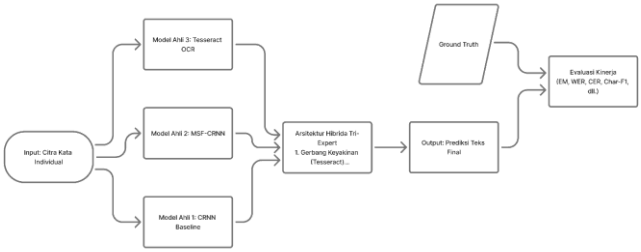
Penelitian Terkait

Riset OCR modern menunjukkan dua kecenderungan utama. Pertama, sistem praktis seperti PP-OCR menekankan efisiensi *pipeline* agar tetap dapat digunakan pada perangkat terbatas (Du et al., 2020). Kedua, *recognizer* berbasis *transformer* dan *language* model seperti ABINet, TrOCR, PARSeq, MATRN, dan VisionLAN meningkatkan akurasi melalui *attention, pretraining*, atau koreksi kontekstual (Fang et al., 2021; Li et al., 2021; Na et al., 2022; Bautista & Atienza, 2022; Wang et al., 2021).

Dibandingkan arah penelitian tersebut, kontribusi penelitian ini berada pada desain *system-level* yang menggabungkan *recognizer* ringan dan OCR klasik melalui mekanisme keputusan terkalibrasi. Pendekatan ini tidak dimaksudkan untuk menggantikan model *transformer* dalam semua konteks, tetapi menawarkan solusi pragmatis untuk *receipt* OCR yang memproses banyak token dan membutuhkan hasil yang mudah diaudit.

3. METODE PENELITIAN

Penelitian ini menggunakan desain eksperimen komparatif pada tahap pengenalan teks level kata. Deteksi area teks tidak menjadi fokus utama karena dataset CORD menyediakan anotasi token dan *bounding box*. Dengan demikian, evaluasi diarahkan untuk mengukur kualitas *word recognition* dan efektivitas strategi fusi *Tri-Expert* secara lebih terkontrol. Diagram alir metodologi penelitian dapat dilihat pada Gambar 1.



Gambar 1. Diagram alir metodologi penelitian

Dataset dan Praproses

Dataset yang digunakan adalah *Consolidated Receipt Dataset* (CORD), yaitu dataset struk publik yang menyediakan citra struk dan anotasi token/kata. Pada implementasi penelitian, anotasi token diekstraksi menjadi *word-crops*, kemudian indeks data disusun ke dalam *split train*, *validation*, dan *test*. Pendekatan *word-crop* dipilih agar perbandingan antar *recognizer* tidak bias oleh kesalahan tahap deteksi.

Praproses pada jalur *deep learning* meliputi konversi citra menjadi *grayscale*, *resize* dengan tinggi tetap H=64, normalisasi piksel, serta *padding* atau pemotongan lebar sesuai konfigurasi pelatihan. Pada jalur *Tesseract*, word-*crop* dapat di-*upscale* sebelum OCR untuk membantu keterbacaan karakter kecil. Setelah inferensi, keluaran semua model dinormalisasi, termasuk pemangkasan spasi dan penyelarasan format karakter agar evaluasi konsisten.

Tabel 1. Konfigurasi eksperimen dan komponen system

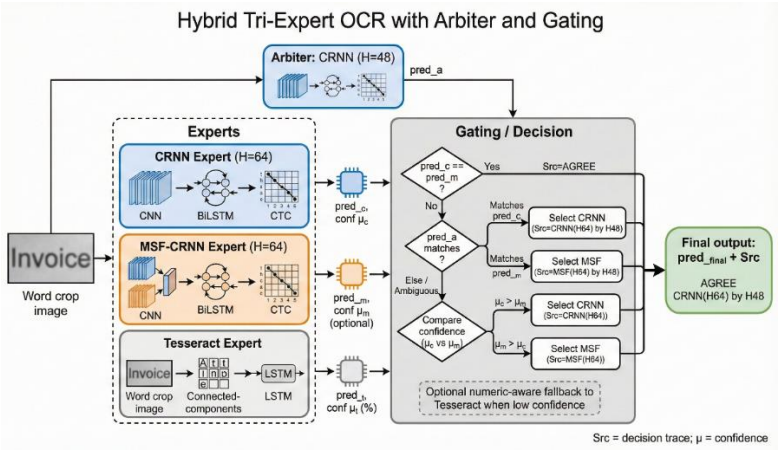
Komponen	Konfigurasi Ringkas
Dataset	CORD, *word-crops* dari anotasi token dan *bounding box*
Split evaluasi	Validation N=2.186 dan Test N=2.356
Expert 1	CRNN(H=64), CNN + BiLSTMx2 + CTC *greedy decoding*
Expert 2	MSF-CRNN(H=64), *multi-stream feature fusion* + CTC
Expert 3	*Tesseract* OCR, OEM 1, PSM 8, bahasa eng+ind
Arbiter	CRNN(H=48) sebagai *tie-breaker* berbasis aturan
Metrik	EM, WER, CER, *Char-Precision, Char-Recall, Char*-F1

Sumber: Hasil pengolahan eksperimen pada dataset CORD.

Arsitektur Hibrida Tri-Expert

Setiap *word-crop* diproses secara paralel oleh CRNN(H=64), MSF-CRNN(H=64), dan *Tesseract*. Hasil utama sistem tidak ditentukan oleh voting sederhana, tetapi melalui dua tahap keputusan. Tahap pertama memilih kandidat antara CRNN(H=64) dan MSF-CRNN(H=64) berdasarkan *confidence* yang telah dikalibrasi dengan *temperature scaling*. Prediksi MSF dipilih hanya apabila selisih *confidence* terhadap CRNN melewati ambang margin tertentu; jika tidak, sistem mempertahankan prediksi CRNN.

Tahap kedua menggunakan CRNN(H=48) sebagai arbiter ketika prediksi CRNN(H=64) dan MSF-CRNN(H=64) berbeda. Jika prediksi arbiter sama dengan salah satu kandidat utama, kandidat tersebut dipilih. Jika tidak ada kecocokan, keputusan kembali mengikuti hasil seleksi tahap pertama. Selain itu, *numeric-aware fallback* berbasis *Tesseract* diaktifkan secara kondisional pada token numerik berkepercayaan rendah, terutama ketika perbedaan prediksi berkaitan dengan digit, pemisah ribuan, desimal, atau spasi. Arsitektur model hibrida Tri-Expert dengan arbiter dan gating dapat dilihat pada Gambar 2.



Gambar 2. Arsitektur model hibrida *Tri-Expert* dengan *arbiter* dan gating

Metode Evaluasi

Kinerja model dievaluasi pada data *validation* dan *test* menggunakan empat metrik utama. EM menunjukkan persentase prediksi yang sama persis dengan *ground truth*. WER pada penelitian ini dihitung pada unit *word-crop* sehingga nilainya ekuivalen dengan 1 - EM. CER dihitung berdasarkan rasio jarak edit karakter terhadap panjang *ground truth.* Char-F1 digunakan untuk menilai keseimbangan presisi dan recall pada level karakter. Pelaporan dua *split* dilakukan untuk melihat konsistensi performa, bukan hanya hasil pada satu data evaluasi.

4. HASIL DAN PEMBAHASAN

Eksperimen menunjukkan bahwa *Hybrid Tri-Exper*t memberikan peningkatan yang konsisten dibanding CRNN(H=64) dan *Tesseract*. Peningkatan terhadap CRNN memang tidak sebesar peningkatan terhadap *Tesseract* karena CRNN sudah menjadi *baseline* kuat pada *word-crop* yang relatif bersih. Namun, kenaikan yang berulang pada *validation* dan *test* menunjukkan bahwa mekanisme fusi memberikan kontribusi nyata, terutama pada token pendek, angka, dan simbol.

Tabel 2. Metrik validasi CRNN, *Tesseract*, dan *Hybrid Tri-Expert*

Model	N	EM (%)	WER (%)	CER (%)	*Char*-P (%)	*Char*-R (%)	*Char*-F1 (%)
CRNN	2.186	90,90	9,10	2,57	98,14	97,52	97,83
Tesseract	2.186	74,47	25,53	9,05	92,93	91,60	92,26
Hybrid	2.186	91,49	8,51	2,37	98,21	97,70	97,96

Sumber: Hasil pengolahan eksperimen pada dataset CORD

Pada data *validation*, *Hybrid Tri-Expert* mencapai EM 91,49%, WER 8,51%, CER 2,37%, dan Char-F1 97,96%. Dibanding CRNN(H=64), metode hibrida meningkatkan EM sebesar +0,59 poin dan menurunkan CER sebesar -0,20 poin. Dibanding *Tesseract*, selisihnya jauh lebih besar, yaitu +17,02 poin EM dan penurunan CER sebesar -6,68 poin. Hasil ini mengindikasikan bahwa *Tesseract* kurang stabil sebagai *recognizer* tunggal pada *word-crops* CORD, tetapi tetap bermanfaat sebagai *fallback* selektif untuk token tertentu.

Tabel 3. Metrik pengujian CRNN, *Tesseract*, dan *Hybrid Tri-Expert*

Model	N	EM (%)	WER (%)	CER (%)	Char-P (%)	Char-R (%)	Char-F1 (%)
CRNN	2.356	88,12	11,88	3,47	97,31	96,74	97,02
Tesseract	2.356	68,97	31,03	12,11	90,46	88,78	89,60
Hybrid	2.356	89,52	10,48	3,13	97,43	97,06	97,25

Sumber: Hasil pengolahan eksperimen pada dataset CORD.

Pada data *test*, pola peningkatan tetap konsisten. *Hybrid Tri-Expert* mencapai EM 89,52%, WER 10,48%, CER 3,13%, dan Char-F1 97,25%. Dibanding CRNN(H=64), metode hibrida meningkatkan EM sebesar +1,40 poin dan menurunkan CER sebesar -0,33 poin. Dibanding *Tesseract*, metode hibrida unggul +20,55 poin EM dan menurunkan CER sebesar -8,98 poin. Konsistensi ini memperkuat dugaan bahwa fusi *multi-expert* bukan hanya memperbaiki satu *split* tertentu, melainkan membantu menangani *failure mode* yang muncul berulang pada *receipt* OCR.

Secara konseptual, peningkatan berasal dari tiga sumber. Pertama, MSF-CRNN membantu ketika informasi stroke tipis atau detail lokal lebih penting dibanding representasi CRNN standar. Kedua, *confidence calibration* mengurangi risiko sistem memilih model yang *overconfident* tetapi salah. Ketiga, *numeric-aware fallback* memberi jalur koreksi untuk token

angka dan format harga yang sensitif terhadap kesalahan satu karakter. Kombinasi ini menjadikan *Hybrid Tri-Expert* lebih mudah diaudit karena keputusan setiap token dapat dilacak melalui *expert* yang dipilih, *confidence*, arbiter, dan pemanggilan *fallback*.

Jika dibandingkan dengan pendekatan *transformer* modern, metode ini tidak mengklaim menjadi solusi dengan kapasitas terbesar. *Transformer* seperti TrOCR, PARSeq, ABINet, MATRN, dan VisionLAN berpotensi menghasilkan akurasi lebih tinggi pada *benchmark* tertentu, tetapi biaya komputasinya lebih besar dan integrasinya lebih kompleks. Untuk skenario struk yang memproses banyak *word-crops*, rancangan hibrida ini menawarkan kompromi yang lebih praktis: tetap ringan, modular, dapat dipadukan dengan sistem *Tesseract* yang sudah ada, dan memiliki jejak keputusan yang lebih jelas untuk *debugging*.

Implementasi demo berbasis Gradio dan *HuggingFace Spaces* menunjukkan bahwa *pipeline* dapat dijalankan di luar lingkungan eksperimen. Demo tersebut memuat alur unggah citra struk, deteksi atau pemrosesan kata, pengenalan berbasis *Tri-Expert*, rekonstruksi teks, dan ringkasan keluaran. Aspek ini penting karena kontribusi penelitian tidak berhenti pada metrik, tetapi juga menunjukkan kesiapan awal untuk diuji secara praktis pada aplikasi OCR struk.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil merancang dan mengimplementasikan OCR hibrida *Tri-Expert* untuk pengenalan teks level kata pada citra struk. Sistem menggabungkan CRNN(H=64), MSF-CRNN(H=64), dan *Tesseract* melalui seleksi *confidence* terkalibrasi, arbiter CRNN(H=48), serta *numeric-aware fallback*. Berdasarkan evaluasi pada dataset CORD, metode hibrida secara konsisten melampaui *baseline* CRNN(H=64) dan *Tesseract*. Pada *validation*, metode hibrida mencapai EM 91,49%, CER 2,37%, dan Char-F1 97,96%; pada test, metode hibrida mencapai EM 89,52%, CER 3,13%, dan Char-F1 97,25%. Peningkatan terhadap CRNN memang relatif moderat, tetapi konsisten dan relevan karena terjadi pada token yang rentan menimbulkan kesalahan praktis seperti angka, simbol, dan kata pendek.

Hasil penelitian menunjukkan bahwa pendekatan hibrida lebih tepat dipahami sebagai strategi peningkatan reliabilitas daripada sekadar penggantian *recognizer*. Keunggulannya terletak pada modularitas, *auditability*, dan kemampuan memilih *expert* yang lebih sesuai untuk setiap token. Dengan demikian, sistem lebih mudah diintegrasikan ke *pipeline* OCR yang telah ada dan lebih mudah dianalisis ketika terjadi kesalahan. Keterbatasan penelitian ini adalah evaluasi utama masih berfokus pada *word-crop* dengan deteksi teks yang diasumsikan tersedia,

sehingga belum sepenuhnya menggambarkan tantangan *end-to-end* pada halaman struk penuh. Penelitian berikutnya disarankan menguji sistem pada skenario *full-page receipt*, mengoptimalkan inferensi *batch*, menambahkan validasi pasca-OCR berbasis struktur transaksi, serta mengeksplorasi integrasi *recognizer* berbasis *transformer* seperti TrOCR, PARSeq, atau ABINet secara selektif agar peningkatan akurasi tetap sebanding dengan biaya komputasi.

UCAPAN TERIMA KASIH

Artikel ini merupakan pengembangan dari penelitian tugas akhir pada Program Studi Informatika, Universitas Muhammadiyah Semarang. Penulis menyampaikan terima kasih kepada dosen pembimbing dan pihak-pihak yang telah memberikan arahan, masukan, serta dukungan selama proses penelitian dan penyusunan naskah.

DAFTAR REFERENSI

A-Sawaareekun, C., & Lipikorn, R. (2025). Menu item extraction from Thai receipt images using deep learning and template-based information extraction. https://doi.org/10.1145/3704391.3704407

Atienza, R. (2021). Vision transformer for fast and efficient scene text recognition. In Proceedings of the International Conference on Document Analysis and Recognition (ICDAR). https://doi.org/10.1109/ICDAR52589.2021.00010

Baek, J., Kim, G., Lee, J., Park, S., Han, D., Yun, S., Oh, S. J., & Lee, H. (2019). What is wrong with scene text recognition model comparisons? Dataset and model analysis. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 4715-4723). https://doi.org/10.1109/ICCV.2019.00481

Bautista, D., & Atienza, R. (2022). Scene text recognition with permuted autoregressive sequence models (PARSeq). In European Conference on Computer Vision (ECCV). http://arxiv.org/abs/2207.06966

Dietterich, T. G. (2000). Ensemble methods in machine learning. In Multiple Classifier Systems (pp. 1-15). Springer.

Du, Y., Li, C., Guo, R., Yin, X., Liu, W., Zhou, J., & Bai, X. (2020). PP-OCR: A practical ultra lightweight OCR system. http://arxiv.org/abs/2009.09941

Fang, S., Xie, H., Wang, Y., Mao, Z., & Zhang, Y. (2021). Read like humans: Autonomous, bidirectional and iterative language modeling for scene text recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 7098-7107).

Graves, A., Fernandez, S., Gomez, F., & Schmidhuber, J. (2006). Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In Proceedings of ICML 2006 (pp. 369-376).

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In Proceedings of ICML 2017 (pp. 1321-1330).

http://arxiv.org/abs/1706.04599

Hegghammer, T. (2022). OCR with Tesseract, Amazon Textract, and Google Document AI: A benchmarking experiment. Journal of Computational Social Science, 5(1), 861-882. https://doi.org/10.1007/s42001-021-00149-1

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, Z., Chen, K., He, J., Bai, X., Karatzas, D., Lu, S., & Jawahar, C. V. (2019). ICDAR2019 competition on scanned receipt OCR and information extraction. Proceedings of the International Conference on Document Analysis and Recognition, 1516-1520. https://doi.org/10.1109/ICDAR.2019.00244

Ismail, M., Abu Mangshor, N. N., Fadzil, A., & Ibrahim, S. (2024). Automated receipt scanning using convolutional recurrent neural network (CRNN). https://doi.org/10.1109/SCOReD64708.2024.10872626

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. Neural Computation, 3(1), 79-87. https://doi.org/10.1162/neco.1991.3.1.79

Lestari, I. N. T., & Mulyana, D. I. (2022). Implementation of OCR (Optical Character Recognition) using Tesseract in detecting character in quotes text images. Journal of Applied Engineering and Technological Science, 4(1), 58-63. https://doi.org/10.37385/jaets.v4i1.905

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady, 10(8), 707-710.

Li, M., Lv, T., Cui, L., Lu, Y., Florencio, D. A. F., Zhang, C., Li, Z., & Wei, F. (2021). TrOCR: Transformer-based optical character recognition with pre-trained models. http://arxiv.org/abs/2109.10282

Lin, C. J., Liu, Y. C., & Lee, C. L. (2022). Automatic receipt recognition system based on artificial intelligence technology. Applied Sciences, 12(2). https://doi.org/10.3390/app12020853

Na, B., Kim, Y., & Park, S. (2022). Multi-modal text recognition networks: Interactive enhancements between visual and semantic features (MATRN). In European Conference on Computer Vision (ECCV).

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62-66. https://doi.org/10.1109/TSMC.1979.4310076

Park, S., Shin, S., Lee, B., Lee, J., Surh, J., Seo, M., & Lee, H. (2019). CORD: A consolidated receipt dataset for post-OCR parsing. In Document Intelligence Workshop at Neural Information Processing Systems (NeurIPS).

Sauvola, J., & Pietikainen, M. (2000). Adaptive document image binarization. Pattern Recognition, 33(2), 225-236. https://doi.org/10.1016/S0031-3203(99)00055-2

Sayallar, C., Sayar, A., & Babalik, N. (2023). An OCR engine for printed receipt images using deep learning techniques. International Journal of Advanced Computer Science and Applications, 14. https://doi.org/10.14569/IJACSA.2023.0140295

Shi, B., Bai, X., & Yao, C. (2017). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Transactions

on Pattern Analysis and Machine Intelligence, 39(11), 2298-2304. https://doi.org/10.1109/TPAMI.2016.2646371

Smith, R. (2007). An overview of the Tesseract OCR engine. Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), 2, 629-633. https://doi.org/10.1109/ICDAR.2007.4376991

Wang, P., Jin, L., Zhang, Y., Zhu, C., Shen, C., & Cao, W. (2021). VisionLAN: Visual language model for scene text recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). http://arxiv.org/abs/2108.09661

Wang, X., Zhang, X., Lei, S., & Deng, H. (2020). A method of text detection and recognition from receipt images based on CRAFT and CRNN. Journal of Physics: Conference Series, 1518, 012053. https://doi.org/10.1088/1742-6596/1518/1/012053

Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241-259. https://doi.org/10.1016/S0893-6080(05)80023-1

Yu, W., Ibrayim, M., & Hamdulla, A. (2023). Scene text recognition based on improved CRNN. Information, 14(7), 1-14. https://doi.org/10.3390/info14070369

Zou, L., He, Z., Wang, K., Wu, Z., Wang, Y., Zhang, G., & Wang, X. (2023). Text recognition model based on multi-scale fusion CRNN. Sensors, 23(16). https://doi.org/10.3390/s23167034