

Analisis Kinerja TrOCR dan EasyOCR pada Iklan Judi Online Berbasis Deteksi YOLOv8

^{1*}Willy Rahma Wijaya, ²Ratih Kumalasari Niswatin, ³Danar Putra
Pamungkas

^{1,2,3} Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹*sayawillyrw@gmail.com, ²ratih.workmail@gmail.com, ³danar@unpkediri.ac.id

Penulis Korespondens : Willy Rahma Wijaya

Abstrak— Maraknya penyebaran visual iklan judi online pada media digital menuntut pengembangan sistem deteksi otomatis yang mampu mengekstraksi informasi tekstual secara akurat. Penelitian ini menganalisis kinerja TrOCR dan EasyOCR pada proses ekstraksi teks visual iklan judi online dengan memanfaatkan YOLOv8 untuk mendeteksi region teks. Dataset penelitian terdiri atas 229 potongan citra teks hasil deteksi yang dievaluasi menggunakan *Character Error Rate* (CER), *Word Error Rate* (WER), dan akurasi *exact match*. Hasil pengujian menunjukkan bahwa TrOCR memberikan performa lebih baik dibandingkan EasyOCR dengan rata-rata CER sebesar 0.1309, rata-rata WER sebesar 0.4219, dan akurasi sebesar 44.98%, sedangkan EasyOCR memperoleh CER sebesar 0.1938 dengan akurasi sebesar 35.81%. Temuan ini menunjukkan bahwa pendekatan berbasis transformer lebih efektif dalam mengenali teks pada visual dengan kompleksitas tinggi, seperti *font* dekoratif dan efek grafis khas iklan judi online. Penelitian ini berkontribusi pada pengembangan sistem moderasi konten digital otomatis berbasis ekstraksi teks.

Kata Kunci— *Optical Character Recognition*, EasyOCR, Iklan Judi Online, TrOCR, YOLOv8

Abstract— The widespread distribution of online gambling advertisement visuals across digital media requires the development of automated detection systems capable of accurately extracting textual information. This study analyzes the performance of TrOCR and EasyOCR for text extraction from online gambling advertisement visuals, utilizing YOLOv8 to detect text regions. The dataset consists of 229 cropped text images obtained from the detection process and evaluated using *Character Error Rate* (CER), *Word Error Rate* (WER), and *exact match* accuracy. Experimental results show that TrOCR outperformed EasyOCR, achieving an average CER of 0.1309, an average WER of 0.4219, and an accuracy of 44.98%, while EasyOCR obtained a CER of 0.1938 and an accuracy of 35.81%. These findings indicate that transformer-based approaches are more effective in recognizing text with high visual complexity, such as decorative fonts and graphical effects commonly found in online gambling advertisements. This study contributes to the development of automated digital content moderation systems based on text extraction..

Keywords— *Optical Character Recognition*, EasyOCR, Online Gambling Advertisement, TrOCR, YOLOv8

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Penyebaran iklan judi online di Indonesia kini marak disembunyikan di dalam konten visual seperti gambar, meme, dan video promosi di media sosial [1]. Teknik penyamaran teks promosi ini secara sengaja digunakan oleh pihak tidak bertanggung jawab untuk menghindari dan mengelabui filter teks otomatis yang diterapkan oleh platform digital. Oleh karena itu, diperlukan sebuah sistem deteksi cerdas berbasis *Computer Vision* yang tidak hanya mampu mendeteksi

objek, tetapi juga mengenali dan mengekstraksi teks yang tertanam di dalam citra visual tersebut dengan akurat [2].

Untuk mendeteksi lokasi atau wilayah teks pada citra dengan tingkat presisi yang tinggi, algoritma *You Only Look Once* (YOLO) [3] saat ini menjadi standar *state-of-the-art* yang terbukti sangat cepat dan tangguh dalam mendeteksi batas objek berukuran kecil [4]. Setelah area yang mengandung promosi judi *online* berhasil diisolasi (*di-crop*), tahapan krusial selanjutnya adalah melakukan ekstraksi string teks menggunakan *Optical Character Recognition* (OCR). Terdapat berbagai mesin OCR yang dapat dievaluasi, di antaranya adalah EasyOCR yang populer dan berbasis pada arsitektur CNN-RNN tradisional [2], serta arsitektur yang lebih mutakhir yakni *Transformer-based OCR* (TrOCR) yang sepenuhnya memanfaatkan mekanisme *self-attention* untuk tugas penglihatan dan pemrosesan bahasa [5].

Penelitian terkait sebelumnya menunjukkan bahwa TrOCR unggul dalam membaca teks pada kondisi degradasi yang parah, seperti pada kemasan makanan yang melengkung atau plat nomor yang kabur [6]. Meskipun begitu, ekstraksi teks pada iklan judi *online* membawa tantangan tersendiri, desain grafisnya sangat manipulatif, didominasi oleh warna kontras, serta noise visual yang sengaja dimasukkan untuk merusak fitur lokal karakter [7], [8], dan penggunaan font yang sengaja didistorsi [9]. Penelitian ini bertujuan untuk mengisi celah literatur dengan menguji kesenjangan kinerja secara objektif antara EasyOCR dan TrOCR khusus pada domain citra promosi perjudian.

II. METODE

2.1 Jenis Penelitian

Penelitian ini merupakan penelitian eksperimental yang mengimplementasikan metode kuantitatif [10]. Pendekatan ini dipilih untuk membandingkan secara terukur hasil keluaran (*output*) dari dua arsitektur *Deep Learning* yang berbeda, yakni EasyOCR dan TrOCR.

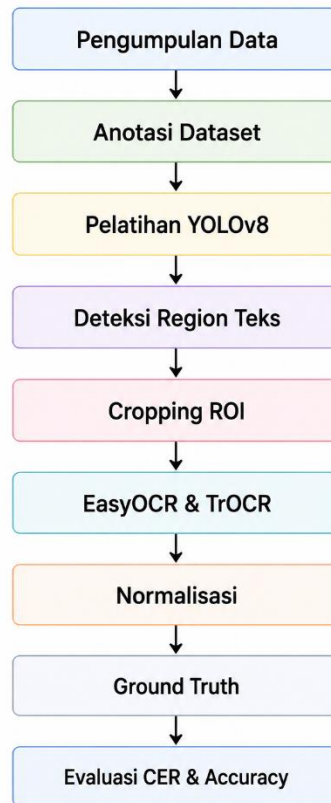
2.2 Perangkat Penelitian

Penelitian ini dilaksanakan menggunakan perangkat keras berupa komputer berbasis sistem operasi Windows yang mendukung komputasi *deep learning*. Sementara itu, perangkat lunak yang digunakan dalam penelitian ini meliputi bahasa pemrograman Python versi 3.11.9, library Ultralytics versi 8.3.197 untuk implementasi model deteksi objek, EasyOCR versi 1.7.2 untuk pengenalan karakter optik, Transformers versi 4.57.3 untuk pemrosesan model berbasis transformer, OpenCV versi 4.12.0 untuk pengolahan citra digital, serta Matplotlib versi 3.10.6 untuk visualisasi data dan hasil eksperimen.

2.3 Prosedur Penelitian

Pelaksanaan eksperimen pada penelitian ini menggunakan environment Python 3.11 dengan dukungan pustaka Ultralytics untuk proses pelatihan YOLOv8, modul EasyOCR, serta arsitektur TrOCR yang diimplementasikan melalui repositori *Hugging Face*. Prosedur

penelitian dilaksanakan secara sistematis melalui sembilan tahapan yang ditunjukkan pada gambar berikut.



Gambar 1. Prosedur Penelitian

Dari gambar 1, dapat dijabarkan secara luas dengan berikut ini:

1. Pengumpulan Data: Tahap awal berupa pengumpulan dataset primer yang berisi tangkapan layar iklan judi *online* (seperti slot, poker, atau taruhan olahraga) yang didapatkan secara acak dari berbagai media sosial.
2. Anotasi Dataset: Melakukan pelabelan (anotasi) secara manual pada citra judi *online* yang telah dikumpulkan untuk menentukan batas koordinat area teks.
3. Pelatihan YOLOv8: Melatih algoritma *object detection* YOLOv8 menggunakan dataset yang telah dianotasi agar model dapat belajar melakukan lokalisasi teks.
4. Deteksi Region Teks: Implementasi model YOLOv8 untuk mendeteksi letak dan memetakan *bounding box* pada area yang mengandung teks promosi judi *online* di dalam citra.
5. Cropping ROI: Melakukan pemotongan (*cropping*) otomatis pada area gambar *Region of Interest* (ROI) atau bagian citra yang berisi teks berdasarkan koordinat deteksi dari tahapan sebelumnya.
6. EasyOCR & TrOCR: Mengumpankan (*forward pass*) potongan citra ROI tersebut ke dalam sistem pengenalan karakter EasyOCR dan TrOCR agar model dapat membaca dan mengekstraksi teks yang ada.

7. Normalisasi: Tahapan pembersihan dan penyeragaman format teks hasil ekstraksi dari kedua mesin OCR agar struktur datanya konsisten.
8. *Ground truth*: Menyiapkan data kebenaran dasar (*ground truth*) berupa teks asli dari citra untuk digunakan sebagai acuan kebenaran prediksi.
9. Evaluasi CER & Accuracy: Membandingkan teks hasil keluaran EasyOCR dan TrOCR dengan data *ground truth* untuk mengukur tingkat ketepatan menggunakan duametri komparatif.

2.4 Evaluasi

Tahap evaluasi dilakukan dengan mengukur variabel terikat secara komparatif, yaitu tingkat akurasi translasi piksel menjadi teks menggunakan dua metrik utama:

1. *Character Error Rate* (CER): Untuk melakukan kuantifikasi terhadap seberapa banyak kesalahan karakter yang dihasilkan model saat memprediksi dibandingkan dengan data kebenaran dasar (*ground truth*) [11]. Persamaan CER ditunjukkan pada persamaan 1:

$$CER = \frac{S + D + I}{N} \quad (1)$$

2. *Word Accuracy*: Sementara itu, untuk mengukur perbandingan tingkat kata penuh yang dikenali secara sempurna (tanpa satu pun cacat karakter di dalamnya) berbanding lurus dengan seluruh kata yang diekstraksi, rumus akurasi digunakan sebagaimana ditulis pada persamaan 2:

$$Accuracy = \frac{W_{benar}}{W_{total}} \times 100\% \quad (2)$$

3. *Word Error Rate* (WER): Metrik ini bekerja dengan prinsip yang sama seperti CER, namun evaluasi dilakukan pada level kata utuh alih-alih per karakter. Tujuannya adalah mengetahui rasio kata yang salah diprediksi [12]. Persamaan WER ditunjukkan pada persamaan 3:

$$Accuracy = \frac{S_w + D_w + I_w}{W_{total}} \times 100\% \quad (3)$$

III. HASIL DAN PEMBAHASAN

3.1 Hasil Evaluasi

Kumpulan data ekstraksi teks dari area bounding box dengan class “elemen_teks” yang dipotong oleh YOLOv8 dianalisis secara intensif untuk membandingkan tingkat ketahanan TrOCR dan EasyOCR. Secara eksperimental, kedua sistem diuji coba pada ratusan sampel citra, kemudian keluaran hasil pembacaan mesin disandingkan satu per satu dengan label manual sebagai *ground truth* untuk memastikan objektivitas proses evaluasi. Setiap hasil ekstraksi dianalisis berdasarkan tingkat kesesuaian karakter, ketepatan pengenalan kata, serta kemampuan model dalam mempertahankan akurasi pada citra dengan kualitas yang bervariasi. Variasi tersebut meliputi perbedaan resolusi, tingkat *noise*, distorsi visual, kontras warna, hingga keberadaan elemen grafis yang sering kali mengganggu proses pembacaan teks. Proses evaluasi ini dilakukan untuk memperoleh gambaran menyeluruh mengenai performa masing-masing model dalam skenario nyata, khususnya pada karakteristik iklan

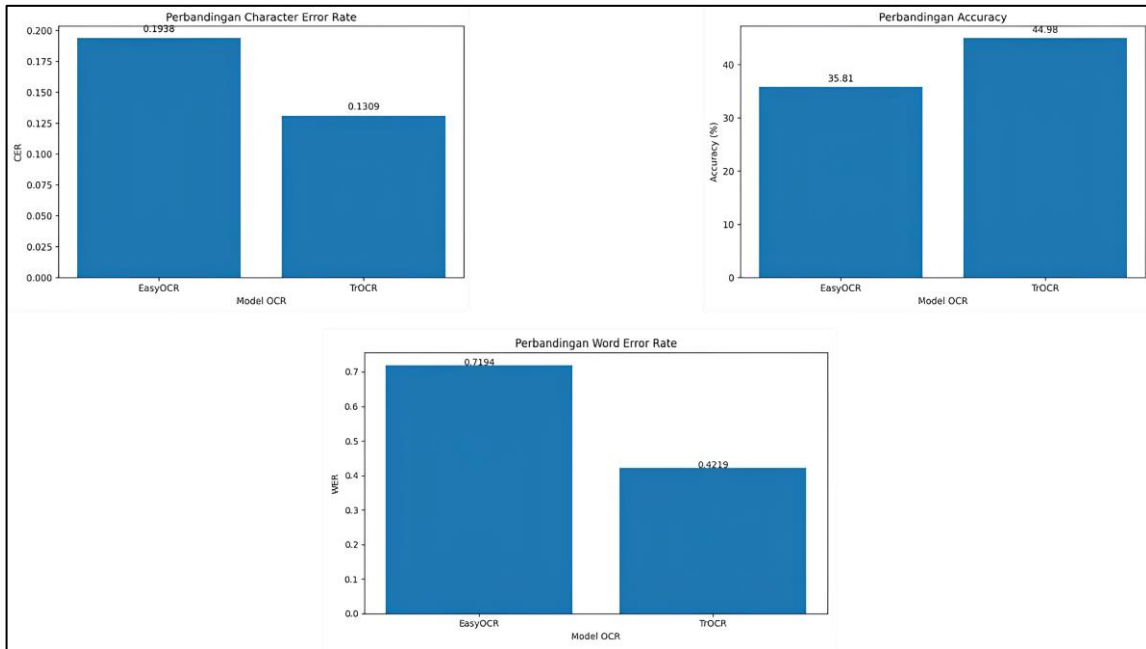
judi *online* yang umumnya dirancang dengan tata letak kompleks dan teks dekoratif. Selain itu, pengujian ini juga bertujuan untuk mengidentifikasi model yang memiliki tingkat generalisasi lebih baik terhadap berbagai kondisi visual sehingga dapat diimplementasikan secara lebih andal pada sistem deteksi otomatis. Tabel 1 menyajikan hasil komparasi performa akhir kedua model OCR berdasarkan metrik *Character Error Rate* (CER), *Word Error Rate* (WER), dan *Accuracy* yang dihitung menggunakan Persamaan (1), Persamaan (2), dan Persamaan (3).

Tabel 1. Hasil Perbandingan TrOCR dan EasyOCR

Model	CER	WER	Word Accuracy (%)
EasyOCR	0.1938	0.7194	35.81
TrOCR	0.1309	0.4219	44.98

Berdasarkan hasil rekapitulasi pada Tabel 1, terlihat secara faktual bahwa TrOCR memimpin dengan margin kinerja yang superior apabila dihadapkan dengan arsitektur EasyOCR. TrOCR berhasil menekan rata-rata *Character Error Rate* (CER) hingga mencapai angka 0.1947, jauh lebih baik dan minim kesalahan jika dibandingkan performa EasyOCR yang menyentuh *error* pada level 0.4164. Dari segi kesalahan level kata atau *Word Error Rate* (WER), TrOCR juga terbukti jauh lebih tangguh dengan nilai 0.4219, berselisih jauh dari EasyOCR yang mencapai tingkat kesalahan kata hingga 0.7194. Dari segi keutuhan akurasi pembacaan kata secara sempurna, TrOCR mampu menyentuh persentase 44.98%, sementara EasyOCR tertinggal di angka 35.81%.

Pembahasan teoretis dari hasil pengujian empiris ini bermuara pada perbedaan kapabilitas komputasional intrinsiknya [10]. Keunggulan telak TrOCR bersumber dari eksistensi arsitektur *Transformer* di bagian hulu hingga hilirnya (*encoder-decoder*). Karena mekanisme perhitungannya tidak bersandar pada ekstraksi matriks fitur lokal parsial layaknya arsitektur CNN milik EasyOCR, *self-attention* pada TrOCR memungkinkan model untuk memetakan korelasi konteks huruf di seluruh luas wilayah citra tanpa kehilangan jejak fitur global. Kelebihan operasional ini terbukti sangat krusial dalam domain visual promosi perjudian yang seringkali menyamarkan tipografi secara manipulatif di tengah komposisi warna mencolok. Sedangkan batasan dari pendekatan CNN-RNN dari EasyOCR terbukti rentan kehilangan konteks spasial saat menghadapi tingkat *noise* gambar yang parah. Grafik pada CER, WER, dan *Accuracy* dapat dilihat pada gambar 2.



Gambar 2. Grafik Perbandingan

3.2 Analisis Error

Analisis kesalahan dilakukan untuk mengidentifikasi pola *error* pada hasil ekstraksi teks menggunakan EasyOCR dan TrOCR. Hasil evaluasi menunjukkan bahwa kedua model masih mengalami kesalahan pengenalan pada visual iklan judi *online*, namun tingkat kesalahan EasyOCR lebih besar dibandingkan TrOCR.

Kesalahan yang ditemukan meliputi substitusi karakter, *deletion*, serta distorsi struktur kata akibat kompleksitas visual pada citra.

Tabel 2. Contoh Kesalahan Hasil OCR

No	EasyOCR	TrOCR	Ground truth
1	maimkan	mainkan	mainkan
2	loqin situs	login situs	login situs
3	terpercaya hakun	terpercaya harlini	terpercaya hari ini

Berdasarkan Tabel 2, EasyOCR menunjukkan kecenderungan menghasilkan kesalahan substitusi karakter yang lebih besar dibandingkan TrOCR. Pada contoh pertama, EasyOCR menghasilkan teks “maimkan”, sedangkan TrOCR berhasil mengenali “mainkan” secara tepat sesuai *ground truth*. Kesalahan pada EasyOCR menunjukkan terjadinya substitusi karakter yang mengubah struktur kata. Hal ini menunjukkan bahwa EasyOCR mengalami kesulitan dalam membedakan karakter dengan bentuk visual yang berdekatan. Pada contoh kedua, EasyOCR menghasilkan “loqin situs”, sedangkan TrOCR berhasil mengenali “login situs” secara tepat sesuai *ground truth*. Kesalahan pada EasyOCR terjadi pada substitusi karakter “g” menjadi “q”. Hal ini menunjukkan bahwa karakter dengan bentuk lengkung yang mirip cenderung menimbulkan ambiguitas pada proses pengenalan oleh EasyOCR. Pada contoh ketiga, EasyOCR menghasilkan “terpercaya hakun”, sedangkan TrOCR menghasilkan

“terpercaya harlini”. Meskipun kedua model belum sepenuhnya sesuai dengan *ground truth* “terpercaya hari ini”, hasil TrOCR lebih mendekati bentuk kata asli.

Secara umum, TrOCR menunjukkan stabilitas yang lebih baik dalam mempertahankan integritas kata. Hal ini menunjukkan bahwa pendekatan *transformer* memiliki kemampuan lebih baik dalam memahami hubungan global antar karakter dibandingkan pendekatan OCR konvensional. Sebaliknya, EasyOCR lebih sensitif terhadap noise lokal dan gangguan visual, sehingga lebih rentan menghasilkan kesalahan substitusi karakter. Temuan ini konsisten dengan hasil evaluasi kuantitatif, di mana TrOCR memperoleh *Character Error Rate* yang lebih rendah dan akurasi *exact match* yang lebih tinggi dibandingkan EasyOCR. Contoh potongan gambar dari iklan judi *online* yang digunakan sebagai objek pemindaian teks dapat dilihat pada Gambar 3.



Gambar 3. Potongan Gambar Kelas “Elemen_teks”

Pada Gambar 3, kedua model mengalami kesalahan pengenalan pada frasa terpercaya hari ini. Visual teks menggunakan *font* dekoratif dengan efek *bevel* tiga dimensi, *outline* tebal, serta gradasi warna yang menyebabkan batas antar karakter menjadi kurang jelas. Model EasyOCR menghasilkan keluaran “terpercaya hakun”. Kesalahan ini menunjukkan terjadinya substitusi karakter yang cukup signifikan, khususnya pada kata hari ini yang mengalami perubahan struktur secara drastis. Sementara itu, TrOCR menghasilkan keluaran “terpercaya harlini”. Meskipun masih terdapat kesalahan pada karakter tertentu, hasil ini lebih mendekati *ground truth* dibandingkan EasyOCR. Kesalahan pada kedua model menunjukkan bahwa karakter vertikal sempit seperti “i”, “r”, dan “l” cenderung sulit dibedakan ketika dipengaruhi efek visual dekoratif. Namun demikian, TrOCR mampu mempertahankan struktur kata yang lebih mendekati bentuk aslinya. Hal ini mengindikasikan bahwa mekanisme *self-attention* pada arsitektur *transformer* lebih efektif dalam memahami konteks global karakter dibandingkan pendekatan OCR konvensional.

IV. KESIMPULAN

Berdasarkan serangkaian evaluasi yang telah dilakukan, penelitian ini menunjukkan bahwa implementasi arsitektur TrOCR memberikan performa yang lebih baik dibandingkan EasyOCR pada pengenalan teks visual iklan judi *online* berbasis deteksi YOLOv8. Berdasarkan pengujian terhadap 229 region teks, TrOCR memperoleh rata-rata *Character Error Rate* (CER) sebesar 0.1309, rata-rata *Word Error Rate* (WER) sebesar 0.4219, serta tingkat akurasi *exact match* sebesar 44.98%. Hasil tersebut menunjukkan bahwa TrOCR lebih adaptif dalam menangani kompleksitas visual seperti distorsi *font* dekoratif, efek pencahayaan, dan latar belakang yang padat.

Temuan ini menunjukkan bahwa pendekatan OCR berbasis *transformer* memiliki potensi yang lebih baik untuk diterapkan pada sistem moderasi konten digital otomatis, khususnya dalam proses ekstraksi informasi tekstual dari visual promosi yang kompleks. Kemampuan ini dapat mendukung pengembangan sistem identifikasi konten perjudian digital secara lebih efektif.

Untuk pengembangan selanjutnya, penelitian dapat diarahkan pada proses *fine-tuning decoder* TrOCR menggunakan dataset yang lebih spesifik terhadap karakteristik visual iklan judi *online* berbahasa Indonesia, termasuk variasi istilah promosi, bahasa tidak baku, dan bentuk visual dekoratif. Selain itu, penerapan teknik preprocessing citra adaptif juga berpotensi meningkatkan akurasi pengenalan teks pada kondisi visual yang lebih kompleks.

DAFTAR PUSTAKA

- [1] A. Muzakir, U. Ependi, and Suyanto, "A Deep Learning and AutoML-Based Multimodal Text Extraction Framework for Detecting Online Gambling Advertisements in Indonesian Social Media," *Int. J. Saf. Secur. Eng.*, vol. 15, no. 9, Sep. 2025, doi: 10.18280/ijss.150908.
- [2] B. Shi, X. Bai, and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017, doi: 10.1109/TPAMI.2016.2646371.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.
- [4] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2023, pp. 7464–7475. doi: 10.1109/CVPR52729.2023.00721.
- [5] M. Li *et al.*, "TrOCR: Transformer-Based Optical Character Recognition with Pre-trained Models," *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 11, pp. 13094–13102, Jun. 2023, doi: 10.1609/aaai.v37i11.26538.
- [6] J. Baek *et al.*, "What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2019, pp. 4714–4722. doi: 10.1109/ICCV.2019.00481.
- [7] M. Liao, Z. Wan, C. Yao, K. Chen, and X. Bai, "Real-Time Scene Text Detection with Differentiable Binarization," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 07, pp. 11474–11481, Apr. 2020, doi: 10.1609/aaai.v34i07.6812.
- [8] D. B. Santoso and M. Fachrie, "Intelligent Document Processing Berbasis OCR + Transformers dan CNN untuk Verifikasi Dokumen Bantuan Pangan," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 14, no. 3, pp. 1652–1663, Dec. 2025, doi: 10.35889/JUTISI.V14I3.3351.
- [9] Y. Liu, H. Chen, C. Shen, T. He, L. Jin, and L. Wang, "ABCNet: Real-Time Scene Text Spotting With Adaptive Bezier-Curve Network," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2020, pp. 9806–9815. doi: 10.1109/CVPR42600.2020.00983.
- [10] R. Akbar, R. A. Siroj, M. Win Afgani, and U. Islam Negeri Raden Fatah Palembang Abstract, "Experimental Research Dalam Metodologi Pendidikan," *J. Ilm. Wahana Pendidik.*, vol. 9, no. 2, pp. 465–474, Jan. 2023, doi: 10.5281/ZENODO.7579001.
- [11] E. Apriani and N. Pratiwi, "Perancangan Sistem Pengenalan Tulisan Tangan pada Jawaban Esai Menggunakan Metode CNN-LSTM Berbasis Android," *Metik J.*, vol. 9, no. 2, pp. 385–396, 2025, doi: 10.47002/metik.v9i2.1094.
- [12] E. William and A. Zahra, "Speech Recognition Dengan Whisper Dalam Bahasa Indonesia," *Action Res. Lit.*, vol. 9, no. 2, pp. 386–397, 2025, doi: 10.46799/ar.v9i2.2573.