

Klasifikasi Tulisan Tangan Bahasa Arab Menggunakan Teknik Yolo

Muhammad Ilham Zuhurf Firdaus
Fakultas Teknik Elektro
Telkom University
Bandung, Indonesia

Sofia Saidah
Fakultas Teknik Elektro
Telkom University
Bandung, Indonesia

Nur Ibrahim
Fakultas Teknik Elektro
Telkom University
Bandung, Indonesia

Abstrak — Tulisan tangan dalam bahasa Arab memiliki tantangan tersendiri dalam proses pengenalan otomatis karena bentuk hurufnya yang kompleks, bersambung, serta adanya variasi gaya penulisan individu. Permasalahan ini menjadi penting untuk dipecahkan terutama dalam konteks digitalisasi dokumen dan aplikasi pendidikan berbasis teknologi. Penelitian ini bertujuan untuk mengembangkan sistem pendeteksi tulisan tangan bahasa Arab dengan memanfaatkan teknik You Only Look Once (YOLO), sebuah pendekatan object detection berbasis deep learning yang dikenal karena kecepatannya dalam mengenali objek secara real-time. Penelitian ini dibatasi pada deteksi huruf-huruf tunggal dalam tulisan tangan, bukan pada pengenalan kata utuh atau kalimat.

Metode yang digunakan melibatkan beberapa tahap utama, dimulai dari pengumpulan dataset tulisan tangan bahasa Arab yang telah dilabeli secara manual. Selanjutnya dilakukan proses pelatihan model menggunakan arsitektur YOLOv9 dengan menyesuaikan parameter agar optimal terhadap karakteristik huruf Arab. Model dilatih untuk mendeteksi dan mengklasifikasikan huruf secara individu dalam gambar tulisan tangan. Setelah pelatihan, dilakukan evaluasi menggunakan metrik seperti mean Average Precision (mAP), presisi, dan recall untuk mengukur performa sistem. Dalam proses ini juga dilakukan augmentasi data untuk meningkatkan generalisasi model terhadap berbagai bentuk tulisan tangan.

Hasil dari penelitian ini menunjukkan bahwa penggunaan YOLO efektif dalam mendeteksi tulisan tangan huruf Arab dengan akurasi yang cukup tinggi, terutama dalam kondisi pencahayaan dan latar belakang yang bervariasi. Sistem yang dikembangkan mampu melakukan deteksi secara cepat dan akurat, serta memiliki potensi untuk dikembangkan lebih lanjut dalam aplikasi OCR (Optical Character Recognition) untuk bahasa Arab. Penelitian ini diharapkan dapat menjadi langkah awal dalam pengembangan sistem pengenalan tulisan tangan Arab yang lebih kompleks di masa depan.

Kata kunci— YOLO, Tulis tangan, OCR, Bahasa Arab

I. PENDAHULUAN

Bahasa Arab merupakan salah satu bahasa tertua di dunia, kaya akan sejarah dan pengetahuan. Banyak arkeolog yang masih bekerja keras untuk mengungkap fakta-fakta

menariknya, mengingat akarnya yang dalam yang dapat ditelusuri kembali ke abad keenam. Sementara beberapa orang, khususnya mereka yang beragama Islam, mungkin telah mengenal bahasa Arab sejak kecil, orang lain yang baru mengenal bahasa tersebut mungkin merasa sistem penulisannya agak asing.

Tulisan tangan sulit untuk disimpan dan diakses secara efisien dan sesuai. Proses mengidentifikasi dan mengenali berbagai cara pada input seperti gambar, ucapan atau aliran karakter disebut pattern recognition (PR). Tahapan dalam proses tersebut dapat melibatkan pengukuran objek- untuk mengidentifikasi atribut yang membedakan, ekstraksi fitur untuk atribut yang menentukan dan perbandingan dengan pola yang diketahui untuk menentukan apakah ada kecocokan antara keduanya atau tidak (ketidakcocokan).

Pengenalan karakter tulisan tangan Arab tetap menjadi masalah yang sulit karena kekhususan bahasa Arab. Alfabet Arab terdiri dari 28 huruf tanpa huruf besar dan kecil. Setiap huruf memiliki dua atau empat bentuk, tergantung pada posisinya dalam kata, yang mungkin berada di awal, tengah, akhir, dan terisolasi. Dalam alfabet Arab, 15 huruf dari 28 huruf memiliki satu titik atau lebih. Beberapa bentuk karakter serupa, tetapi perbedaan muncul dengan posisi dan jumlah titik yang berada di atas atau di bawah karakter menurut penelitian sebelumnya dengan judul “Investigation on Deep Learning for Off-line Handwritten Arabic Character Recognition” [1].

Meskipun sejumlah bahasa telah menerima perhatian yang signifikan dalam penelitian OCR. Optical Character Recognition (OCR) adalah teknologi yang dapat mengubah tulisan yang dicetak, ditulis tangan, atau diketik menjadi teks digital yang dapat dibaca mesin. Bahasa Arab- masih sebagian besar merupakan bidang studi yang belum banyak dieksplorasi, dengan sebagian besar pekerjaan sebelumnya pada bahasa ini berbasis karakter cetak. Pada penelitian “Arabic Handwritten Characters Recognition using Convolutional Neural Network” menggunakan metode Convolutional Neural Network (CNN) [2]. Convolutional Neural Network (CNN) adalah jenis khusus dari multilayer feed-forward yang dilatih dalam mode terawasi. CNN dilatih dan diuji menggunakan basis data yang berisi 16.800 karakter

tulisan tangan Arab. Dalam makalah tersebut, metode optimasi diimplementasikan untuk meningkatkan kinerja CNN. Metode pembelajaran mesin umum biasanya menggunakan kombinasi antara ekstraktor fitur dan pengklasifikasi yang dapat dilatih. Penggunaan CNN menghasilkan peningkatan signifikan di berbagai pendekatan pembelajaran mesin.

Metode lain yang digunakan adalah penggunaan teknik *congealing* untuk memodelkan distorsi nyata pada tulisan tangan dan kemudian menggunakan model distorsi ini untuk mensintesis contoh tulisan tangan [3]. Adapun *Discrete Cosine Transform (DCT)* adalah teknik untuk mengonversi nilai pixel gambar menjadi komponen frekuensi dasarnya [4]. Dengan menerapkan DCT pada setiap gambar karakter, koefisien DCT dari gambar tersebut dapat diperoleh. Koefisien bernilai tinggi terkumpul di sudut kiri atas, sementara koefisien bernilai rendah berada di sudut kanan bawah array dan *Discrete Wavelet Transform (DWT)* adalah teknik lain yang digunakan untuk mengekstraksi fitur karakter, di mana pada setiap tingkat dekomposisi filter *low-pass (LPF)* dan filter *high-pass (HPF)* diterapkan pada -setiap baris/kolom gambar untuk mendekomposisi gambar menjadi satu sub-pita frekuensi rendah (LL) dan tiga sub-pita frekuensi tinggi (LH, HL, HH) [5].

Dalam penelitian yang berjudul “Multi-Font Farsi/Arabic Isolated Character Recognition Using Chain Codes”, menerapkan menggunakan *chain code-based*. Diawali dengan menipiskan karakter tulisan, fitur vektor yang terdiri dari berbagai atribut dari setiap karakter Farsi/Arab diekstraksi. Bagian paling efisien dari fitur vektor karakter ini adalah *chain code* baru, yang dihitung menggunakan *neighborhood function* dan serangkaian normalisasi. Hasil eksperimen yang didapatkan menggunakan beberapa font Farsi standar menunjukkan bahwa metode yang diusulkan memiliki akurasi rata-rata 97,4% dalam pengenalan karakter yang benar [6].

Penelitian ini mengusulkan pendekatan OCR pada tulisan bahasa Arab menggunakan YOLO. YOLO adalah singkatan dari istilah 'You Only Look Once'. ini sering dikaitkan dengan aturan yang mendeteksi dan mengenali banyak objek dalam sebuah gambar (secara *real-time*). YOLOv9 disesuaikan dan dilatih untuk mengenali bahasa Arab. Namun, YOLO tidak menghasilkan tingkat pengenalan karena satu karakter mungkin memiliki beberapa hasil pengenalan. Penelitian ini meningkatkan OCR YOLOv9 dengan menggunakan tiga teknik: Menghapus hasil pengenalan yang berlebihan dengan memproses kotak yang tumpang tindih tumpang tindih dan menjaga kotak dengan akurasi tertinggi akurasi tertinggi. Selain itu, pemeriksa ejaan dengan jarak edit digunakan untuk mengoreksi kata-kata yang salah. Pada penelitian sebelumnya yang berjudul “Comprehensive Performance Evaluation of YOLOv11, YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environment” [7] membahas perbandingan performansi dari YOLO versi 8 sampai dengan versi 11, dimana YOLOv9 *Gelan-base* and YOLOv9 *Gelan-e* memiliki akurasi konfigurasi tertinggi dan

YOLOv11 memiliki keunggulan dalam validasi penghitungan.kajian teori

Menyajikan dan menjelaskan teori-teori yang berkaitan dengan variabel-variabel penelitian. Poin subjudul ditulis dalam abjad.

A. Model YOLOv9

YOLOv9 merupakan arsitektur penglihatan komputer yang mendukung mendeteksi objek dan segmentasi gambar. YOLOv9 -dikembangkan berdasarkan pendahulunya, mengintegrasikan metode yang mengatasi kendala informasi sambil mendorong batas akurasi dan efisiensi dalam mendeteksi sebuah objek. Berbeda dengan pendahulunya YOLOv8 yang berfokus- pada optimasi kecepatan dan akurasi menggunakan CSPNet dan PANet yang ditingkatkan, YOLOv9 memperkenalkan dua arsitektur baru, yaitu: *Programmable Gradient Information (PGI)* dan *Generalized Efficient Layer Aggregation-Network (GELAN)*. Hal ini bertujuan untuk mengatasi masalah kritis informasi yang hilang saat data melewati lapisan-lapisan jaringan, sehingga meningkatkan tingkat stabilitas gradien dan akurasi prediksi.

Melalui integrasi PGI, YOLOv9 sangat efektif mengatasi masalah kehilangan data dalam jaringan dalam, dan memastikan fitur-fitur penting dan pembangkitan gradien yang handal agar hasil pelatihan optimal. Penggunaan GELAN menghasilkan arsitektur jaringan yang memaksimalkan manfaat dari parameter dan efisiensi komputasi, menjadikan YOLOv9 ini menjadi adaptif dan berkinerja tinggi di berbagai aplikasi. Model ini menunjukkan hasil peningkatan yang luar biasa dalam hal efisiensi, akurasi, dan adaptabilitas, dan menetapkan tolak ukur baru pada dataset MS COCO. Meskipun proyek YOLOv9 dikembangkan oleh tim *open-source* yang berbeda, YOLOv9 dibangun di atas basis kode yang kuat dari *Ultralytics YOLOv5*, mencerminkan semangat kolaboratif komunitas riset AI. Dalam upaya mencapai deteksi objek yang optimal, YOLOv9 menonjol dengan pendekatan yang inovatif untuk mengatasi tantangan hilangnya informasi yang melekat pada jaringan saraf dalam. Dengan mengintegrasikan PGI dan GELAN yang serbaguna, YOLOv9 tidak hanya meningkatkan kapasitas pembelajaran model, tetapi juga memastikan informasi penting selama proses deteksi, sehingga mencapai akurasi dan kinerja yang luar biasa [9].

B. Karakteristik Bahasa Arab

Penulisan bahasa Arab dimulai dari kanan ke kiri dengan total 28 karakter dasar dalam konstruksi alfabetnya. Bahasa Arab juga terdiri dari enam belas huruf Arab unik yang terdiri dari satu hingga tiga titik. Jumlah dan posisi titik-titik ini membuat perbedaan antara karakter-karakter ini, jika tidak, karakter-karakter ini mirip (seperti خ, ج, dan ح). Selain itu, beberapa karakter (seperti ك, ي, و, ا) memiliki goresan seperti zig-zag (*Hamzah ء*). Titik-titik dan *Hamzah* ini sering disebut sebagai titik sekunder dan dapat ditemukan di atas bagian utama karakter seperti pada kasus ALEF (ا), atau di bawah seperti pada kasus BAA (ب), atau di tengah seperti pada kasus JEEM (ج) [8].

II. METODE

Mengumpulkan tulisan tangan Arab berbagai huruf dari berbagai penulis pada sebuah kertas. Kemudian difoto sehingga mendapatkan gambar tulisan tersebut. Gambar kertas yang telah didapat kemudian di proses berupa upscaling dan grayscaling. Selanjutnya dibagi untuk training dan testing serta diberi anotasi sesuai dengan huruf. Ada 28 jenis huruf yang digunakan, mulai dari Alif (ا) sampai Ya (ي) dan termasuk huruf vokalnya. Setelah dataset disiapkan, proses training dilakukan menggunakan Kaggle. Hasil training berupa sebuah model (output weight) dalam bentuk meta file dan pb file yang akan digunakan dalam mendeteksi huruf pada testing sets. Tiap hasil deteksi huruf digabungkan sehingga kata pada gambar dapat dikenali. Dari hasil deteksi maupun pembacaan dihitung akurasi dari model yang telah dihasilkan.

III. HASIL DAN PEMBAHASAN

Pengujian perbandingan pada YOLOv9 melibatkan evaluasi kinerja model yang telah dilatih dan diverifikasi. Model yang digunakan yaitu: Model YOLOv9t, Model YOLOv9s, Model YOLOv9m, Model YOLOv9c, dan Model YOLOv9e. Tabel berikut ini memberikan ringkasan tentang varian model YOLOv9, termasuk ukuran gambar standar 640 piksel, jumlah mAPval 50-95, mAPval 50, parameter, dan FLOPs.

TABEL 1
Hasil Pengujian Model Sistem YoloV9

Model	size (pixels)	mAP ^{val} 50-95	mAP ^{val} 50	params (M)	FLOPs (B)
YOLOv9t	640	0.0128	0.0374	2.0	7.6
YOLOv9s	640	0.0713	0.168	7.2	26.8
YOLOv9m	640	0.267	0.491	20.0	76.6
YOLOv9c	640	0.359	0.604	25.3	102.4
YOLOv9e	640	0.404	0.65	57.4	189.3

IV. KESIMPULAN

Pada penelitian ini telah dibuat dataset tulisan arab berupa huruf hijaiyah dengan total 28 huruf dan total 542 dataset dengan jumlah masin-masing hurufnya kurang lebih 19 huruf, data tersebut dianotasikan dengan roboflow sehingga diperoleh dataset yang sudah diberi kelas, kemudian dataset

tersebut dilatih menggunakan YOLOv9 dan diuji sehingga diperoleh hasil terbaik mAPval 50 60%, persisi 85%, recall 65%, F-Score 16,1%. Hasil tersebut diperoleh ketika menggunakan epochs 30, image size 640 pixel, batch size 8.

REFERENSI

- [1] C. Boufenar, A. Kerboua and M. Batouche, Investigation on Deep Learning for Off-line Handwritten Arabic Character Rec, 2017.
- [2] A. Elsayy, H. M. El-Bakry and M. Loey, "Arabic Handwritten Characters Recognition using Convolutional Neural Network," WSEAS TRANSACTIONS ON COMPUTER RESEARCH, 2017.
- [3] M. Shatnawi and S. Abdallah, "Improving Handwritten Arabic Character Recognition by Modeling," ACM Transactions on Asian and Low-Resource Language Information Processing, 2015.
- [4] A. Lawgali, "An Evaluation of Methods for Arabic Character Recognition," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 7, 2014.
- [5] B. Wang, J. Ding, Q. Wen, X. Liao and C. Liu, AN IMAGE WATERMARKING ALGORITHM BASED ON DWT DCT AND SVD, Beijing, 2009.
- [6] H. I. S. A. M. B. T. L. and K. Z. , "Multi-Font Farsi/Arabic Isolated Character Recognition Using Chain Codes," 2008.
- [7] R. Sapkota, . Z. Meng, M. Churuvija, X. Du, Z. Ma and M. Karkee, "Comprehensive Performance Evaluation of YOLO11, YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environments," Center for Precision & Automated Agricultural Systems, 2024.
- [8] H. . O. K. Sugianto, . M. A. D. Widyadara and A. . B. Setiawan, "IMPLEMENTATION OF FACE RECOGNITION FOR ATTENDANCE USING YOLO V3 METHOD," Seminar Nasional Inovasi Teknologi , Kediri, 2022.
- [9] . G. Huang, Y. Huang, H. Li, Z. Guan, X. Li, G. Zhang, W. Li and . X. Zheng, "An Improved YOLOv9 and Its Applications for Detecting Flexible Circuit Boards Connectors," International Journal of Computational Intelligence Systems , 2024.