

Deteksi Karakter pada Tulisan Tangan Latin Menggunakan You Only Look Once (YOLO)

Aisyah Nurul Hidayah, Yoyon Kusnendar Suprapto

Dept. Teknik Komputer Institut Teknologi Sepuluh Nopember Surabaya, Indonesia

Email: aisyah16@mhs.te.its.ac.id, yoynsuprapto@ee.its.ac.id

Abstrak

Tulisan tangan saat ini masih digunakan dalam berbagai bidang, khususnya tulisan tangan Latin. Namun, terdapat beberapa tantangan dalam mengenali tulisan tangan Latin. Pertama, tulisan tangan memiliki banyak variasi. Variasi yang dimaksud meliputi ukuran huruf, kemiringan huruf, garis dasar atau baseline penulisan, bahkan perbedaan jarak antarhuruf dan antar kata. Kedua, setiap huruf dalam tulisan tangan Latin saling terhubung. Tantangan ini menyebabkan waktu yang lebih lama dalam proses pengenalan tulisan tangan Latin. You Only Look Once (YOLO) digunakan untuk mendeteksi huruf pada tulisan tangan Latin sebagai upaya untuk mempercepat proses pengenalan tulisan tangan Latin. Eksperimen ini dilakukan dengan mengumpulkan sampel citra dari 4 jenis tulisan tangan yang berbeda. Hasil eksperimen menggunakan GPU NVIDIA GeForce RTX 1050 menunjukkan waktu pemrosesan per citra untuk sistem deteksi pada citra satu kata adalah sebesar 0.0776 detik dengan menggunakan nilai ambang (threshold score) sebesar 0.3.

Keyword: *tulisan tangan Latin, You Only Look Once, YOLO*

Diterima Redaksi: 05-Peb-2025 Selesai Revisi: 15-Peb-2025 Diterbitkan Online: 15-Maret-2025
DOI: <https://doi.org/10.59378/jcenim.v3i1.67>

I. LATAR BELAKANG

Tulisan tangan merupakan salah satu bentuk komunikasi. Melalui tulisan tangan, manusia dapat menyampaikan ide tanpa harus berada di tempat yang sama dengan pembaca. Dari hal tersebut, kemudian muncul semakin banyak perkembangan tulisan tangan serta keunikan perkembangan pada setiap budaya di dunia.

Sejarah perkembangan tulisan itu sendiri dimulai dari masyarakat primitif, ketika mereka mulai mengembangkan coretan pada dinding gua, kayu, maupun batu. Seiring berjalananya waktu, coretan-coretan tersebut berkembang menjadi simbol dan pada akhirnya menjadi alfabet atau huruf. Huruf-huruf alfabet kemudian menyebar ke berbagai negara, sehingga memunculkan berbagai jenis alfabet, misalnya alfabet Yunani, alfabet Rusia, dan alfabet Latin. Secara khusus di Eropa, tulisan tangan Latin dikenal sebagai tulisan kursif, yang kemudian juga menyebar ke seluruh dunia.

Tulisan tangan masih digunakan dalam berbagai bidang, salah satunya adalah layanan pengiriman. Menurut Badan Pusat Statistik (BPS), pengiriman surat oleh Pos Indonesia Divisi Regional Jawa Timur pada tahun 2015 mencapai 22.200.517 surat [1] dan paket sebanyak 627.922 [3]. Angka ini diperkirakan akan meningkat pada tahun-tahun berikutnya.

Tulisan tangan, khususnya tulisan tangan Latin, memiliki beberapa tantangan. Pertama, tulisan tangan memiliki variasi yang sangat besar. Variasi tersebut meliputi ukuran huruf, kemiringan huruf dan kata, garis dasar penulisan, tekanan penulisan, bahkan jarak antarhuruf dan antarkata [2]. Kedua, setiap huruf dalam tulisan tangan Latin saling terhubung menjadi satu.

Kedua tantangan tersebut menyebabkan proses pengenalan tulisan tangan Latin membutuhkan waktu yang lebih lama. Selain itu, sistem Optical Character Recognition (OCR) yang ada saat ini belum mendukung pengenalan tulisan tangan Latin. Oleh karena itu, diperlukan sebuah sistem yang dapat mendeteksi huruf pada tulisan tangan Latin sebagai cara untuk mempercepat proses pengenalan. Tugas akhir ini bertujuan untuk membuat sistem deteksi huruf pada tulisan tangan Latin menggunakan You Only Look Once (YOLO). You Only Look Once (YOLO) digunakan karena YOLO mampu mendeteksi objek dengan cepat.

II. PENELITIAN TERKAIT

Beberapa penelitian mengenai tulisan tangan dan pendekripsi objek menggunakan YOLO telah dilakukan oleh beberapa pihak. Penelitian pertama berjudul “Javanese Script Handwriting Mardi Kawi Writing Recognition Rules Become Alphabet Letters On Paper Media Using The Convolutional Neural Network (CNN) Method With You Only Look Once (YOLO) Architecture” [4]. Penelitian ini menggunakan YOLO untuk mendekripsi tulisan tangan dari 5 judul dan penulis yang berbeda pada aksara Jawa Mardikawi. Pengujian menggunakan 33 jenis karakter Jawa Mardikawi menghasilkan nilai rata-rata IoU sebesar 46%, sedangkan pengujian menggunakan 44 jenis karakter Jawa Mardikawi menghasilkan IoU sebesar 40%, yang masing-masing menggunakan dataset dengan ukuran kertas asli.

Penelitian kedua berjudul “TextCaps: Handwritten Character Recognition with Very Small Datasets” [5]. Penelitian ini membahas pengenalan tulisan tangan menggunakan dataset yang relatif kecil. Sistem ini berguna dalam pengenalan karakter untuk bahasa lokal yang tidak memiliki banyak dataset. Dataset tersebut dilakukan augmentasi sehingga diperoleh lebih banyak variasi. Augmentasi ini bersifat realistik dan mencerminkan variasi tulisan tangan manusia yang sebenarnya dengan menambahkan noise terkontrol secara acak. Pengenalan karakter menggunakan arsitektur yang terdiri dari capsule network dan network decoder—hasil yang diperoleh melampaui hasil pengenalan karakter pada dataset EMNIST-letter.

Penelitian ketiga berjudul “Real-Time Face Detection based on YOLO” [6]. Penelitian ini membahas pendekripsi wajah menggunakan jaringan YOLOv3. Pengujian menggunakan dataset WIDER FACE menunjukkan pendekripsi wajah yang lebih cepat dengan waktu deteksi rata-rata sebesar 0,028 detik. Selain lebih cepat, hasil yang diperoleh juga lebih akurat. Kecepatan deteksi tersebut dapat memenuhi kebutuhan waktu nyata dan menghasilkan performa yang baik.

Penelitian keempat berjudul “Thermal Object Detection in Difficult Weather Conditions Using YOLO” [7] yang bertujuan untuk mendekripsi objek pada citra dan video termal yang diambil dalam kondisi cuaca berbeda (cuaca cerah, hujan lebat, dan kabut); kondisi perekaman dilakukan selama pengawasan area yang dilindungi, dengan menggunakan pendekripsi objek waktu nyata. Pendekripsi objek yang digunakan dalam pengujian meliputi Faster RCNN, SSD, FCOS, Cascade RCNN, dan YOLOv3. Berdasarkan hasil pengujian, YOLOv3 memiliki kecepatan pemrosesan tertinggi, yaitu sebesar 27.472 FPS dengan nilai AP sebesar 97,93%.

III. METODOLOGI

Tugas akhir ini merupakan penelitian di bidang visi komputer yang bertujuan untuk mendekripsi huruf pada tulisan tangan Latin menggunakan You Only Look Once (YOLO). Diagram alir pada Gambar 1 merupakan gambaran umum dari sistem yang akan dibuat.

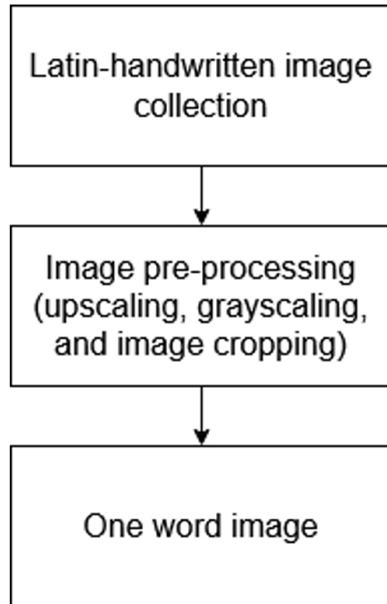


Gambar 1: Rancangan sistem secara keseluruhan

Langkah awal yang dilakukan adalah akuisisi data dengan mengumpulkan berbagai kata tulisan tangan Latin dari berbagai penulis pada selembar kertas. Selanjutnya, kertas tersebut difoto menggunakan kamera sehingga diperoleh citra kertas. Citra kertas yang diperoleh kemudian diproses dalam bentuk peningkatan resolusi dan ukuran (upsampling) serta konversi ke skala abu-abu (grayscale). Citra yang berisi tulisan tangan Latin kemudian dipotong per kata sehingga satu citra hanya berisi satu kata. Kata-kata tersebut dibagi menjadi data latih dan data uji serta dianotasi sesuai dengan hurufnya. Terdapat 26 jenis huruf yang digunakan, mulai dari A hingga Z. Setelah dataset disiapkan, proses pelatihan dilakukan menggunakan Darkflow. Hasil pelatihan berupa sebuah model (bobot keluaran) yang berisi berkas meta dan berkas pb untuk mendekripsi huruf pada data uji. Setiap hasil deteksi huruf kemudian digabungkan sehingga kata pada citra dapat dikenali. Akurasi model dihitung untuk mengevaluasi kinerja model.

A. Akuisisi Data

Pada tugas akhir ini, akuisisi data dilakukan dengan memotret teks pada kertas menggunakan kamera. Pada Gambar 2 dijelaskan tahapan dalam proses akuisisi data.



Gambar 2: Alur akuisisi data

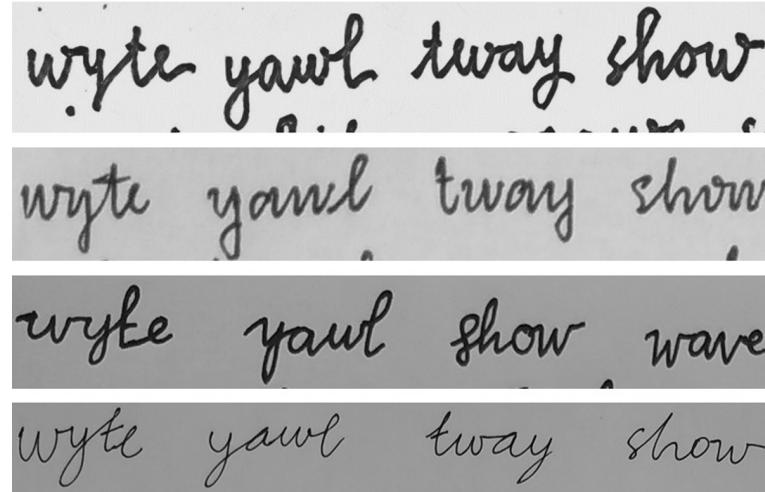
Proses pengumpulan tulisan tangan Latin diawali dengan menyiapkan kombinasi kata yang akan dituliskan. Kombinasi kata yang tepat memengaruhi variasi posisi huruf sehingga setiap huruf dapat berada pada posisi tengah atau akhir kata. Setiap orang menuliskan ulang kombinasi kata yang digunakan sebagai dataset pada selembar kertas. Tulisan tangan Latin dikumpulkan dari beberapa orang untuk memperoleh beberapa lembar kertas dari penulis yang berbeda. Kertas yang berisi tulisan tangan Latin tersebut kemudian diakuisisi dengan cara difoto menggunakan kamera atau dipindai menggunakan pemindai (scanner) printer, sehingga diperoleh citra kertas. Hasil proses pengumpulan citra tulisan tangan ditunjukkan pada Gambar 3.



Gambar 3: Hasil proses pengumpulan citra tulisan tangan

Pra-pemrosesan citra mencakup tiga tahap, yaitu peningkatan kualitas dan ukuran citra (upscale), konversi citra ke skala abu-abu (grayscale), serta pemotongan citra menjadi citra yang hanya berisi satu kata. Citra yang memiliki ukuran di bawah 2000 piksel akan diperbesar sehingga ukurannya menjadi di atas 4000 piksel. Citra kemudian dikonversi ke skala abu-abu setelah proses upscale. Tujuan dari proses grayscale ini adalah untuk menyeragamkan warna tinta yang digunakan pada kertas sehingga

fitur dan latar belakang dapat dibedakan. Proses upscaling dan grayscaling menggunakan perangkat lunak penyunting foto seperti Adobe Photoshop. Hasil proses grayscaling citra ditunjukkan pada Gambar 4.



Gambar 4: Hasil citra setelah proses grayscaling

Citra tulisan tangan Latin kemudian dipotong sehingga satu citra hanya berisi satu kata. Proses pemotongan citra dilakukan menggunakan perangkat lunak Microsoft Photos. Citra kata kemudian dibagi menjadi dua bagian, yaitu data latih dan data uji dengan rasio masing-masing sebesar 70% dan 30%. Hasil pemotongan citra menjadi citra satu kata ditunjukkan pada Gambar 5.



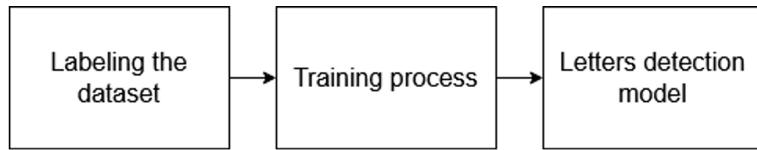
Gambar 5: Hasil pemotongan citra menjadi citra satu kata

B. Pelatihan Data

Setelah proses pengumpulan citra dan pra-pemrosesan citra, tahap selanjutnya adalah pelatihan data. Terdapat beberapa langkah dalam proses ini yang dijelaskan dalam diagram blok pada Gambar 6.

Citra dalam dataset baik data latih maupun data uji kemudian diberikan anotasi atau pelabelan. Pelabelan merupakan proses pemberian label pada citra dengan menentukan bounding box pada objek serta nama kelasnya. Tujuan dari proses pelabelan adalah untuk memperoleh ground-truth bounding box. Proses pelabelan menggunakan LabelImg Tools [8]. Gambar 7 menunjukkan proses pelabelan pada citra satu kata.

Hasil dari proses pelabelan berupa berkas yang berisi informasi koordinat bounding box objek pada citra. Berkas ini disimpan dalam bentuk .xml atau .txt sehingga setiap citra memiliki satu berkas anotasi.



Gambar 6: Alur pelatihan data



Gambar 7: Proses pelabelan pada citra satu kata

Dataset yang diperlukan disiapkan dari proses ini, yang berisi citra dan anotasi citra tersebut. Hasil proses pelabelan ditunjukkan pada Gambar 8.

```

<object>
    <name>w</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
        <xmin>16</xmin>
        <ymin>126</ymin>
        <xmax>76</xmax>
        <ymax>194</ymax>
    </bndbox>
</object>
  
```

Gambar 8: Hasil dari proses pelabelan

Proses pelatihan memerlukan konfigurasi layer dalam bentuk berkas cfg dan bobot pra-latih (pre-trained weight). Selain itu, versi jaringan yang digunakan harus sama dengan versi bobot pra-latih. Pada tugas akhir ini, versi YOLO yang digunakan adalah YOLOv1. Sama seperti pada proses pelabelan, berkas labels.txt pada Darkflow [9] diubah sesuai dengan kelas yang akan digunakan, yaitu huruf Latin a hingga z. Jaringan perlu dikonfigurasi sebelum memulai proses pelatihan. Konfigurasi jaringan disimpan dalam berkas .cfg. Beberapa parameter yang diatur dalam berkas cfg meliputi:

- 1) Mengubah jumlah kelas pada layer region sesuai dengan jumlah kelas yang akan dilatih, nilai width & height, learning rate, serta jumlah filter pada layer konvolusional yang terletak pada dua layer terakhir, sesuai dengan Persamaan 1.

$$\text{filters} = \text{num} \times (\text{classes} + 5) \quad (1)$$

dengan num = 5 dan classes adalah jumlah kelas yang digunakan.

- 2) Menghapus tanda komentar pada baris batch dan subdivisions pada bagian training dengan menghilangkan tanda #.

Rincian pengaturan parameter jaringan untuk tugas akhir ini ditunjukkan pada Tabel 1.

Tabel 1: Rincian Pengaturan Parameter Jaringan

Parameter	Nilai
Width & Height	416x416
Batch & Subdivision	64 & 32
Epoch	1000
Learning rate	0.001
Decay	0.0005
Momentum	0.9
Class	26
Filters	155

Proses pelatihan menggunakan kerangka kerja Darkflow dan Google Colaboratory. Selama proses pelatihan, Darkflow akan menyimpan hasil sementara dalam bentuk TensorFlow checkpoint. Hasil akhir pelatihan berupa bobot akan disimpan dalam berkas .meta dan .pb. Bobot akhir tersebut digunakan sebagai bobot pada model deteksi huruf. Pada pelatihan ini, hasil pelatihan dari checkpoint terakhir disimpan sebagai bobot akhir.

Selain itu, proses pembacaan tulisan tangan juga diperlukan. Pembacaan yang dibahas adalah pembacaan kata. Proses pembacaan berasal dari kelas yang muncul pada hasil prediksi. Kelas-kelas yang muncul kemudian dapat digabungkan sehingga dapat dibaca sebagai sebuah kata. Hasil pembacaan ditampilkan pada sudut kiri atas setiap citra. Algoritma pembacaan dimulai dengan mengambil nilai xmin dan label dari setiap bounding box yang muncul setelah proses prediksi dan menyimpannya ke dalam sebuah array dua dimensi. Isi array tersebut kemudian diurutkan berdasarkan nilai xmin dari setiap bounding box. Nilai xmin diurutkan dari yang terkecil hingga terbesar. Selanjutnya, label akan dikembalikan sesuai dengan urutan pasangan nilai xmin tersebut. Algoritma pembacaan ditunjukkan secara jelas pada diagram alir, seperti pada Gambar 9.

IV. PENGUJIAN DAN ANALISIS

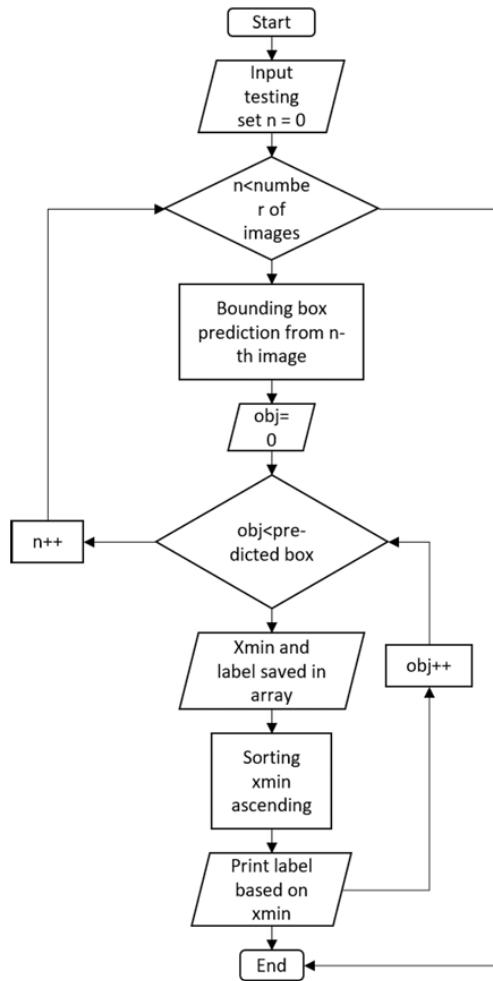
Pada tugas akhir ini, terdapat tiga pengujian model deteksi yang dibagi menjadi tiga bagian sebagai berikut:

- 1) Pengujian deteksi huruf dengan variasi ukuran tulisan tangan
- 2) Pengujian deteksi huruf pada citra satu kata
- 3) Pembacaan tulisan tangan pada citra satu kata

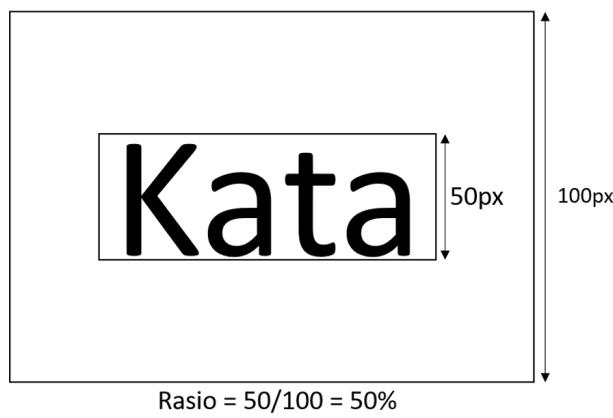
A. Pengujian Deteksi Huruf dengan Variasi Ukuran Tulisan Tangan

Pengujian ini dilakukan untuk mengetahui pengaruh variasi ukuran tulisan terhadap hasil deteksi huruf. Ukuran tulisan tangan merupakan rasio antara tinggi tulisan terhadap tinggi keseluruhan citra dalam satuan piksel. Ilustrasi ukuran tulisan tangan ditunjukkan pada Gambar 10. Pada pengujian ini, terdapat tiga percobaan pelatihan dengan ukuran tulisan tangan yang berbeda. Percobaan 1 merupakan percobaan pelatihan menggunakan dataset dengan rasio 50%. Percobaan 2 merupakan percobaan pelatihan menggunakan dataset dengan rasio 20%. Percobaan 3 merupakan percobaan pelatihan menggunakan dataset gabungan antara percobaan pelatihan 1 dan 2. Hasil dari ketiga percobaan tersebut berupa model yang diuji pada dataset dengan berbagai ukuran tulisan tangan, yaitu rasio 50%, 40%, 20%, dan 10%.

Pada percobaan 1, model tidak dapat mendeteksi ukuran tulisan dengan rasio 10%. Untuk ukuran tulisan dengan rasio 40% dan 20%, model cukup mengalami kesulitan dalam mendeteksi huruf. Meskipun demikian, bounding box yang dihasilkan cukup presisi dalam mendeteksi huruf, terutama pada rasio 50%. Hasil percobaan 1 ditunjukkan pada Gambar 11.



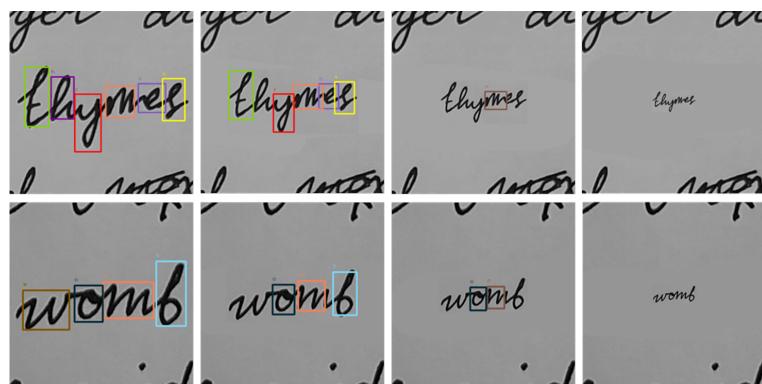
Gambar 9: Diagram alir pembacaan tulisan tangan Latin



Gambar 10: Ilustrasi ukuran tulisan tangan

Pada percobaan 2, terlihat bahwa model dapat mendeteksi huruf pada semua variasi ukuran tulisan. Namun, hasil deteksi pada tulisan dengan rasio 50% tidak sebaik pada tulisan dengan rasio 40%. Bounding box yang dihasilkan hanya presisi pada tulisan dengan rasio 20%. Hasil percobaan 2 ditunjukkan pada Gambar 12.

Pada percobaan 3, terlihat bahwa model tidak dapat mendeteksi huruf pada rasio 10%. Selain itu, banyak bounding box yang tidak akurat atau merupakan false positive. Hal ini terjadi karena model menghasilkan cukup banyak prediksi yang salah. Meskipun demikian, masih terdapat cukup banyak bounding box yang presisi. Hasil percobaan 3 ditunjukkan pada Gambar 13.



Gambar 11: Hasil percobaan 1



Gambar 12: Hasil percobaan 2



Gambar 13: Hasil percobaan 3

B. Pengujian Deteksi Huruf pada Citra Satu Kata

Model yang telah dibuat harus mampu mendeteksi huruf pada citra tulisan. Pengujian ini bertujuan untuk mengetahui hasil deteksi huruf dan waktu pemrosesan pada citra satu kata untuk setiap nilai ambang kepercayaan (threshold confidence score) sehingga dapat diketahui nilai ambang kepercayaan yang optimal. Deteksi dilakukan pada data uji dengan total 56 kata atau 1268 huruf yang dituliskan oleh empat penulis yang berbeda. Hasil deteksi ditunjukkan pada Gambar 14.

Berdasarkan Gambar 14, terdapat sepuluh huruf yang dituliskan oleh penulis. Model telah berhasil mendeteksi huruf pada citra ketika nilai ambang kepercayaan adalah 0.3. Hal ini disebabkan oleh warna tulisan yang sangat kontras dengan latar belakang sehingga huruf dapat terdeteksi dengan baik. Tabel 2 menunjukkan bahwa ketika nilai ambang kepercayaan sama dengan 1, tidak ada kotak prediksi yang muncul. Hal ini terjadi karena tidak ada kotak prediksi yang memiliki nilai kepercayaan sama dengan 1. Nilai ambang akan menghilangkan kotak prediksi apabila kotak prediksi tersebut memiliki nilai



Gambar 14: Hasil deteksi huruf pada citra satu kata (threshold 0.3)

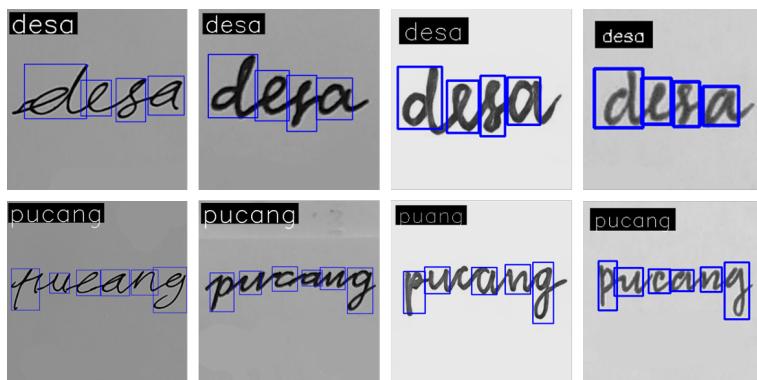
kepercayaan lebih rendah dari nilai ambang yang ditetapkan. Hasil prediksi kemudian dianalisis untuk mengevaluasi kinerja model. Analisis yang digunakan adalah analisis confusion matrix dan mean average precision, seperti ditunjukkan pada Tabel 2.

Tabel 2: Hasil Deteksi dan Pembacaan Huruf pada Citra Satu Kata

Threshold	\hat{B}	Waktu pemrosesan	Confusion Matrix			Kinerja		Kata benar	
			TP	FP	FN	Precision	Recall		
0.1	1463	0.075	1129	334	139	0.70	0.81	84.70	76
0.2	1315	0.080	1107	208	161	0.76	0.79	83.16	110
0.3	1241	0.078	1089	152	179	0.77	0.78	81.93	114
0.4	1167	0.076	1053	114	215	0.79	0.75	79.14	96
0.5	1074	0.077	999	75	269	0.82	0.71	74.68	79
0.6	933	0.075	886	47	382	0.80	0.62	65.81	46
0.7	660	0.077	634	26	634	0.77	0.43	45.97	9
0.8	244	0.075	240	4	1028	0.77	0.15	15.95	0
0.9	33	0.074	33	0	1235	0.27	0.02	2.44	0
1.0	0	0.074	0	0	1268	0.00	0.00	0.00	0

C. Pembacaan Tulisan Tangan pada Citra Satu Kata

Pengujian ini bertujuan untuk mengetahui hasil pembacaan citra. Pembacaan tulisan tangan didasarkan pada hasil prediksi sebelumnya. Pengujian ini menggunakan 56 kata yang dituliskan oleh empat penulis yang berbeda. Hasil pengujian sistem pembacaan pada citra satu kata ditunjukkan pada Tabel 2 dan Gambar 15.



Gambar 15: Hasil pembacaan tulisan tangan pada citra satu kata (threshold 0.3)

Jumlah teks yang terbaca paling banyak terjadi ketika nilai ambang kepercayaan adalah 0.3, yaitu sebanyak 114 kata yang terbaca dengan benar dari total 224 kata. Ketika nilai ambang dinaikkan, jumlah

kata yang terbaca dengan benar mengalami penurunan. Hal ini disebabkan oleh semakin sedikitnya jumlah bounding box yang muncul.

V. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut:

- 1) Berdasarkan hasil pengujian deteksi huruf dengan variasi ukuran tulisan, model pelatihan yang sesuai menggunakan dataset dengan ukuran tinggi tulisan sebesar rasio 50%.
- 2) Berdasarkan hasil pengujian deteksi huruf dan pembacaan pada citra satu kata, nilai ambang kepercayaan (threshold confidence score) yang ideal untuk digunakan adalah sebesar 0.3, dengan waktu pemrosesan per citra sebesar 0.0776 detik, nilai presisi sebesar 76.92%, nilai recall sebesar 78.27%, serta nilai mean average precision (mAP) sebesar 81.93%.
- 3) Sistem pembacaan tulisan tangan telah mampu membaca huruf-huruf yang terdeteksi secara berurutan dari kiri ke kanan.

Daftar Pustaka

- [1] “Surat pos dalam dan luar negeri yang dikirim dan diterima, 2015 (update terakhir 18 Oct 2017),” <https://jatim.bps.go.id/statictable/2017/10/18/701/surat-pos-dalam-dan-luar-negeri-yang-dikirim-dan-diterima-di-provinsi-jawa-timur-2015.html>, 2017, terakhir diakses pada 16 April 2020.
- [2] J. Ryu and H. I. Koo, “Word segmentation method for handwritten documents based on structured learning,” *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1161–1165, 2015.
- [3] “Paket pos dalam dan luar negeri yang dikirim dan diterima, 2015 (update terakhir 18 Oct 2017),” <https://jatim.bps.go.id/statictable/2017/10/18/700/paket-pos-dalam-dan-luar-negeri-yang-dikirim-dan-diterima-di-provinsi-jawa-timur-2015.html>, 2017, terakhir diakses pada 16 April 2020.
- [4] L. E. Damayanti, “Pengenalan tulisan tangan naskah aksara jawa aturan penulisan mardi kawi menjadi huruf alfabet pada media kertas menggunakan metode Convolutional Neural Network (CNN) arsitektur You Only Look Once,” 2020.
- [5] V. Jayasundara, S. Jayasekara, H. Jayasekara, J. Rajasegaran, S. Seneviratne, and R. Rodriguez, “TextCaps: Handwritten character recognition with very small datasets,” *arXiv preprint arXiv:1904.08095*, 2019.
- [6] W. Yang and Z. Jiachun, “Real-time face detection based on YOLO,” in *Proceedings of the 2018 1st International Conference on Control and Communication Engineering*, 2018, pp. 221–224.
- [7] M. Krišto, M. Ivašić-Kos, and M. Pobar, “Thermal object detection in difficult weather conditions using YOLO,” *IEEE Access*, vol. 8, pp. 125459–125476, 2020.
- [8] Tzutalin, “LabelImg,” <https://github.com/tzutalin/labelImg>, 2015, terakhir diakses pada 4 April 2020.
- [9] T. H. Trieu, “Darkflow,” GitHub Repository: <https://github.com/thtrieu/darkflow>, 2018, terakhir diakses pada 18 November 2019.