

# Deteksi Multi-Objek Huruf Hijaiyah Tulisan Tangan Anak Menggunakan YOLO v11s

*Multi-Object Detection of Children's Handwritten Hijaiyah Letters Using YOLOv11s*

Rina Zakiyatun Niza<sup>1</sup>, Akhmad Fadjeri<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Universitas Ma'arif Nahdlatul Ulama Kebumen, Indonesia

E-mail : zaa63450@gmail.com<sup>1</sup>, fadjeri.akhmadfadjeri@gmail.com<sup>2</sup>

\*Corresponding author

Received 25 April 2026; Revised 11 May 2026; Accepted 15 May 2026

**Abstrak** — Masa usia dini merupakan tahap penting dalam pembelajaran membaca Al-Qur'an, khususnya dalam mengenal huruf hijaiyah. Namun, variasi tulisan tangan anak-anak serta kemiripan visual antarhuruf hijaiyah menjadikan proses identifikasi lebih menantang, terutama pada skenario multi-objek dalam satu gambar. Penelitian ini bertujuan untuk mengembangkan sistem deteksi multi-objek secara real-time untuk huruf hijaiyah tulisan tangan anak usia dini menggunakan algoritma YOLOv11s dalam aplikasi berbasis web. Dataset terdiri dari 429 gambar tulisan tangan asli yang memuat lebih dari 3.250 objek huruf beranotasi dalam 20 kelas. Setelah proses preprocessing dan augmentasi data, dataset yang digunakan untuk pelatihan meningkat menjadi 732 gambar, yang dikumpulkan dari siswa usia 4–6 tahun di MDT Ula Addiya Kuwayuhan. Model dilatih menggunakan pendekatan object detection dengan anotasi bounding box dan dievaluasi menggunakan metrik precision, recall, dan mean Average Precision (mAP). Hasil eksperimen menunjukkan bahwa model mencapai nilai precision sebesar 0,867, recall sebesar 0,831, mAP@0.5 sebesar 0,903, dan mAP@0.5:0.95 sebesar 0,700, dengan kecepatan inferensi sekitar 80 FPS. Sistem yang diimplementasikan mampu mendeteksi beberapa huruf hijaiyah secara simultan pada kondisi tulisan tangan yang tidak terstruktur secara real-time. Kontribusi utama penelitian ini terletak pada penerapan YOLOv11s untuk deteksi multi-objek huruf hijaiyah tulisan tangan anak usia dini dengan variasi visual yang tinggi, yang diintegrasikan ke dalam sistem berbasis web secara real-time dan dilengkapi fitur Text-to-Speech (TTS) sebagai media pembelajaran interaktif.

**Kata Kunci** — deteksi objek, huruf hijaiyah, multi-objek, tulisan tangan, pembelajaran interaktif, YOLOv11s.

**Abstract** –Early childhood is a crucial stage in learning to read the Qur'an, particularly in recognizing hijaiyah letters. However, the variability of children's handwriting and the visual similarity among hijaiyah letters make the identification process more challenging, especially in multi-object scenarios within a single image. This study aims to develop a real-time multi-object detection system for handwritten hijaiyah letters of early childhood learners using the YOLOv11s algorithm in a web-based application. The dataset consisted of 429 original handwritten images containing more than 3,250 annotated letter objects across 20 classes. After preprocessing and data augmentation, the dataset used for training increased to 732 images, collected from students aged 4–6 years at MDT Ula Addiya Kuwayuhan. The model was trained using an object detection approach with bounding box annotations and evaluated using precision, recall, and mean Average Precision (mAP) metrics. The experimental results showed that the model achieved a precision of 0.867, recall of 0.831, mAP@0.5 of 0.903, and mAP@0.5:0.95 of 0.700, with an inference speed of approximately 80 FPS. The implemented system was able to detect multiple hijaiyah letters simultaneously under unstructured handwritten conditions in real time. The main contribution of this study lies in the application of YOLOv11s for multi-object handwritten hijaiyah detection in early childhood handwriting with high visual variation, integrated into a

*real-time web-based system equipped with a Text-to-Speech (TTS) feature as an interactive learning medium.*

**Keywords** - *object detection, hijaiyah letters, multi-object, handwriting, interactive learning, YOLOv11s.*

## 1. PENDAHULUAN

Pendidikan Anak Usia Dini (PAUD) merupakan fase krusial dalam perkembangan anak yang sering disebut sebagai masa (*golden age*), dimana kemampuan kognitif, bahasa, dan literasi berkembang secara pesat [1]. Pada tahap ini, pemberian stimulasi yang tepat sangat diperlukan untuk membantu anak dalam mengenali simbol serta mengembangkan kemampuan membaca dan menulis sebagai dasar bagi pembelajaran selanjutnya. Dalam konteks pendidikan islam, pengenalan huruf hijaiyah sejak dini memiliki peran yang sangat penting, karena huruf hijaiyah menjadi dasar utama dalam proses membaca Al-Qur'an[2]. Pemahaman yang baik terhadap bentuk dan pelafalan huruf hijaiyah akan mempermudah anak dalam melanjutkan pembelajaran membaca Al-Qur'an pada tahap selanjutnya [3].

Madrasah Diniyah Takmiliah Ula Addiya Kuwayuhan (MDT Ula Addiya Kuwayuhan) sebagai lembaga pendidikan islam nonformal memiliki santri dengan rentang usia yang beragam, yaitu antara 3 hingga 15 tahun. Namun, fokus penelitian ini berada pada kelompok anak usia dini, khususnya usia 4 hingga 6 tahun, yang menjadi sumber utama dalam proses pengumpulan data. Pembelajaran pada anak usia ini masih menerapkan metode pembelajaran *konvensional* dalam pengenalan huruf hijaiyah, seperti penggunaan buku *Iqro* dan kartu huruf. Berdasarkan hasil *observasi* awal pada tanggal 28 Januari 2026, ditemukan bahwa anak-anak sering mengalami kesulitan dalam membedakan huruf hijaiyah yang memiliki kemiripan bentuk visual, seperti huruf *ba* (ب), *ta* (ت), dan *tsa* (ث) yang hanya dibedakan oleh jumlah serta posisi titik. Selain itu, ketidakkonsistenan tulisan tangan anak juga menyulitkan guru dalam melakukan penilaian, proses evaluasi yang masih dilakukan secara manual membuat kegiatan pembelajaran menjadi kurang efisien dan memerlukan waktu yang lebih lama. Kondisi ini menunjukkan perlunya suatu sistem yang dapat membantu mengidentifikasi tulisan huruf hijaiyah secara otomatis guna mendukung pembelajaran yang lebih efektif[4]. Tulisan tangan anak usia dini memiliki karakteristik yang cukup beragam sehingga menimbulkan tantangan tersendiri dalam penerapan *computer vision*[5]. Ketidakstabilan bentuk huruf, ukuran tulisan yang tidak seragam, serta posisi titik huruf hijaiyah yang sering berubah menyebabkan proses pengenaalan karakter menjadi lebih kompleks [6]. Selain itu, keberadaan beberapa huruf dalam satu citra membuat sistem harus mampu mendeteksi dan menentukan lokasi tiap objek secara akurat. Dengan demikian, deteksi huruf hijaiyah tulisan tangan termasuk ke dalam permasalahan *multi-object detection* yang membutuhkan kemampuan klasifikasi dan lokalisasi objek secara simultan.

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*) dan *machine learning* telah membawa kemajuan signifikan dalam bidang pengolahan citra digital, khususnya dalam pengenalan objek secara otomatis pada suatu citra [7]. Penerapan teknik *computer vision* memungkinkan sistem untuk mengekstraksi informasi visual serta mengidentifikasi objek secara efisien dan akurat [8]. Dalam konteks tulisan tangan anak usia dini, diperlukan metode yang mampu mendeteksi beberapa objek sekaligus dalam satu citra dengan kondisi yang tidak terstruktur. Salah satu pendekatan yang sesuai adalah *You Only Look Once (YOLO)*, yaitu metode *object detection* satu tahap (*single-stage detector*) yang mampu mendeteksi beberapa objek dalam satu proses inferensi dengan kecepatan tinggi [9]. Karakteristik ini menunjukkan bahwa YOLO relevan untuk skenario *multi-object handwritten* hijaiyah yang memerlukan deteksi *real-time* pada beberapa huruf sekaligus dalam satu citra.

Penelitian ini menggunakan YOLOv11s karena memiliki ukuran model yang relatif ringan namun tetap mampu mempertahankan performa deteksi yang baik. YOLOv11s memiliki jumlah

parameter dan kompleksitas komputasi yang lebih rendah sehingga lebih sesuai untuk implementasi sistem berbasis *web* dengan keterbatasan sumber daya komputasi. Selain itu, YOLOv11 mendukung proses ekstraksi fitur multi-skala yang berpotensi membantu model dalam mengenali objek kecil dan detail visual yang halus[10]. Karakteristik ini relevan dengan penelitian ini karena huruf hijaiyah tulisan tangan anak usia dini memiliki variasi bentuk yang tinggi serta beberapa huruf hanya dibedakan oleh titik. Dengan demikian, YOLOv11s dipilih untuk memperoleh keseimbangan antara akurasi dan kecepatan inferensi pada proses deteksi multi-objek secara *real-time*[11]. Model yang dikembangkan kemudian diintegrasikan ke dalam aplikasi berbasis *web* untuk mendukung implementasi deteksi huruf secara *real-time* pada proses pembelajaran. Sistem juga dilengkapi dengan fitur umpan balik suara berbasis *Text-to-Speech* (TTS) sebagai elemen interaktif pendukung pembelajaran. Dengan demikian, penelitian tidak hanya berfokus pada pengembangan model deteksi objek, tetapi juga pada implementasi sistem yang dapat digunakan secara langsung[12].

Beberapa penelitian sebelumnya telah mengkaji pengenalan huruf Arab dan tulisan tangan menggunakan pendekatan kecerdasan buatan. Penelitian oleh [13] menunjukkan bahwa penggunaan YOLOv9 mampu mendeteksi huruf Arab dengan kecepatan dan akurasi yang baik, namun penelitian tersebut belum berfokus pada tulisan tangan anak usia dini yang memiliki variasi visual tinggi. Penelitian lain oleh [14] menggunakan *Convolutional Neural Network* (CNN) dan menghasilkan tingkat akurasi yang tinggi, tetapi terbatas pada klasifikasi satu objek dalam satu citra sehingga kurang sesuai untuk skenario *multi-objek*. Sementara itu, penelitian oleh[15] telah mengintegrasikan YOLO dengan teknologi *Text-To-Speech* (TTS), namun masih berfokus pada huruf alfabet latin sehingga belum dapat diterapkan secara langsung pada pembelajaran huruf hijaiyah

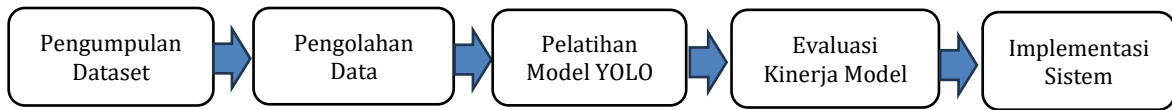
Berdasarkan penelitian terdahulu, masih terdapat keterbatasan pada penerapan deteksi objek untuk tulisan tangan huruf hijaiyah anak usia dini dalam skenario *multi-object* secara *real-time*. Sebagian besar penelitian sebelumnya berfokus pada klasifikasi satu karakter atau menggunakan dataset tulisan tangan yang lebih terstruktur. Padahal, tulisan tangan anak usia dini memiliki tingkat variasi bentuk yang tinggi serta sering mengandung beberapa objek huruf dalam satu citra. Oleh karena itu, penelitian ini difokuskan pada pengembangan sistem *object detection* berbasis YOLOv11s untuk mendeteksi multi-objek huruf hijaiyah tulisan tangan anak usia dini secara *real-time*.

Berdasarkan permasalahan dan kajian penelitian terdahulu, penelitian ini mengembangkan sistem deteksi multi-objek huruf hijaiyah tulisan tangan anak usia dini berbasis YOLOv11s yang mampu melakukan deteksi secara *real-time* pada beberapa huruf dalam satu citra. Sistem menghasilkan *bounding box* dan label kelas pada setiap objek huruf yang terdeteksi, serta dilengkapi dengan fitur *Text-to-Speech* (TTS) sebagai umpan balik interaktif pendukung pembelajaran. Tujuan penelitian ini adalah mengembangkan sistem deteksi *multi-objek* huruf hijaiyah tulisan tangan anak usia dini berbasis YOLOv11s serta menganalisis kemampuan model dalam mengenali huruf dengan kemiripan bentuk visual tinggi pada kondisi tulisan tangan yang tidak terstruktur. Penelitian ini juga mengevaluasi performa model dalam skenario *real-time* berbasis *web* untuk mendukung pembelajaran interaktif huruf hijaiyah.

## 2. METODE PENELITIAN

### 2.1. Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan utama yang meliputi pengumpulan dataset, pengolahan data, pelatihan model YOLO, evaluasi model, serta implementasi sistem berbasis *web* menggunakan *Flask* dan TTS. Alur tahapan penelitian dapat dilihat pada Gambar 1.



**Gambar 1.** Tahapan Penelitian

Tahapan penelitian ini dimulai dari proses pengumpulan dataset tulisan tangan huruf hijaiyah langsung dari siswa MDT Ula Addiya Kuwayuhan. Dataset kemudian diproses melalui tahap *preprocessing*, anotasi *bounding box* dan augmentasi data untuk meningkatkan variasi sampel dan mengurangi *overfitting*. Dataset yang telah diproses selanjutnya digunakan pada tahap pelatihan model YOLOv11s. Model yang dihasilkan kemudian dievaluasi menggunakan metrik *precision*, *recall* dan *mean average Precision (mAP)*. Tahap terakhir adalah implementasi model ke dalam sistem berbasis *web* menggunakan *framework Flask* yang dilengkapi dengan fitur umpan balik suara menggunakan teknologi TTS.

## 2.2. Pengumpulan Dataset

Dataset yang digunakan terdiri dari 429 citra asli tulisan tangan huruf hijaiyah yang dikumpulkan dari siswa usia 4–6 tahun di MDT Ula Addiya Kuwayuhan. Setelah melalui tahap *preprocessing* dan augmentasi data menggunakan Roboflow, jumlah dataset yang digunakan pada proses pelatihan meningkat menjadi 732 citra. Dataset memiliki karakteristik *multi-object* dengan tingkat variasi visual yang tinggi akibat perbedaan bentuk tulisan, ukuran huruf, ketebalan garis, serta posisi titik yang tidak konsisten.

## 2.3. Pengolahan Dataset

### 2.3.1 Preprocessing

Tahap *preprocessing* dilakukan menggunakan platform Roboflow untuk menyiapkan dataset sebelum proses pelatihan model. Dataset diekspor melalui Roboflow dengan *preprocessing* dasar, kemudian pada tahap pelatihan model seluruh citra diproses menggunakan ukuran input  $832 \times 832$  piksel sesuai parameter YOLOv11s. Penggunaan resolusi 832 dipilih untuk membantu model menangkap detail objek kecil, khususnya elemen titik pada huruf hijaiyah yang memiliki kemiripan visual tinggi.

### 2.3.2 Pelabelan Data

Pelabelan data dilakukan menggunakan platform Roboflow dengan teknik *bounding box* untuk menandai lokasi setiap huruf hijaiyah pada citra. Setiap objek diberi label kelas sesuai kategori huruf hijaiyah sehingga menghasilkan dataset terstruktur yang dapat digunakan dalam proses pelatihan model YOLO. Hasil anotasi disimpan dalam format YOLO (.txt) yang memuat informasi kelas objek serta koordinat *bounding box* dalam bentuk  $(x\_center, y\_center, width, height)$  dengan rentang nilai 0–1. Proses anotasi dilakukan secara manual untuk memastikan ketepatan posisi *bounding box* pada setiap objek huruf. Tahap ini menjadi penting karena beberapa huruf memiliki posisi yang berdekatan atau saling tumpang tindih sehingga membutuhkan ketelitian dalam proses pelabelan.

### 2.3.3 Augmentasi Data

Augmentasi data dilakukan untuk meningkatkan variasi dataset, memperbaiki kemampuan generalisasi model terhadap data baru, serta mengurangi risiko *overfitting*. Proses augmentasi dilakukan menggunakan platform Roboflow. Teknik augmentasi yang diterapkan meliputi rotasi ringan ( $\pm 10^\circ$ ), penyesuaian *brightness* dan *exposure*, *gaussian blur*, serta penambahan *noise*. Teknik-teknik ini digunakan untuk mensimulasikan variasi tulisan tangan dan kondisi pencahayaan yang berbeda. Proses augmentasi menghasilkan peningkatan jumlah dataset dari 429 citra asli menjadi 732 citra yang digunakan pada tahap pelatihan model, kemudian secara otomatis

dibagi menjadi data *training*, *validation*, dan *testing*. Selanjutnya, dataset diekspor dalam format YOLOv11 untuk digunakan pada tahap pelatihan model.

#### 2.4. Pelatihan Model YOLO

Proses pelatihan model dilakukan menggunakan *framework* Ultralytics YOLO dengan bahasa pemrograman Python[9]. Model yang digunakan dalam penelitian ini adalah YOLOv11 dengan varian YOLOv11s (*small variant*), yang memiliki jumlah parameter lebih ringan namun tetap mampu memberikan performa deteksi objek yang baik. Dalam proses pelatihan, model melakukan ekstraksi fitur citra melalui jaringan konvolusi untuk mempelajari pola visual setiap huruf hijaiyah. Pelatihan dilakukan pada lingkungan *Google Colab* dengan dukungan GPU Tesla T4 untuk mempercepat proses komputasi.

Untuk kebutuhan pelatihan model, dataset dibagi menjadi data *training*, *validation*, dan *testing* dengan rasio 70:20:10. Pembagian dataset dilakukan secara otomatis menggunakan Roboflow untuk memastikan distribusi data yang proporsional pada setiap subset. Untuk mengurangi potensi *overfitting* akibat jumlah dataset yang relative terbatas, penelitian ini menerapkan data augmentation pada tahap preprocessing serta memantau training loss dan validation loss selama proses pelatihan. Selain itu, *early stopping* digunakan untuk menghentikan proses pelatihan Ketika performa validasi tidak lagi mengalami peningkatan secara signifikan.. Spesifikasi perangkat keras dan perangkat lunak yang digunakan ditunjukkan pada Tabel 1.

**Tabel 1.** Spesifikasi Perangkat Lunak

Komponen	Spesifikasi
Framework	Ultralytics YOLO
Bahasa Pemrograman	Python
GPU	Tesla T4
Data pelatihan	732 hasil augmentasi
Jumlah Kelas	20 Kelas Huruf

Parameter pelatihan yang digunakan dalam penelitian ini ditunjukkan pada Tabel 2.

**Tabel 2.** Parameter Pelatihan Model

Komponen	Spesifikasi
Model	YOLOv11s
Epoch	200
Image Size	832
Batch Size	16

Model dilatih menggunakan *optimizer* bawaan YOLO dengan pendekatan *stochastic gradient descent* (SGD) untuk meminimalkan fungsi *loss* yang terdiri dari *bounding box loss*, *classification loss*, dan *distribution focal loss*. Selama proses pelatihan, model secara iteratif memperbarui bobot jaringan untuk meningkatkan kemampuan dalam mengenali pola visual huruf hijaiyah berdasarkan dataset yang telah dianotasi.

#### 2.5. Evaluasi Model YOLO

Evaluasi kinerja model YOLO dilakukan menggunakan beberapa metrik kuantitatif, yaitu *precision*, *recall*, *F1-score*, dan *mean Average Precision* (mAP) yang umum digunakan dalam sistem deteksi objek[16]. Penggunaan metode *deep learning* dalam berbagai penelitian menunjukkan bahwa pendekatan ini efektif dalam melakukan klasifikasi citra dengan tingkat akurasi yang tinggi [17]. *Precision* digunakan untuk mengukur tingkat ketepatan prediksi model terhadap objek yang terdeteksi, sedangkan *recall* digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh objek yang seharusnya terdeteksi. *F1-score* merupakan rata-rata harmonik antara *precision* dan *recall* yang digunakan untuk mengevaluasi keseimbangan antara kedua metrik tersebut[18].

*Precision* dihitung menggunakan persamaan berikut:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

*Recall* dihitung menggunakan persamaan berikut :

$$Precision = \frac{TP}{TP+FN} \quad (2)$$

*F1-Score* dihitung sebagai rata-rata harmonik antara *precision* dan *recall* dengan persamaan berikut :

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

**Keterangan :**

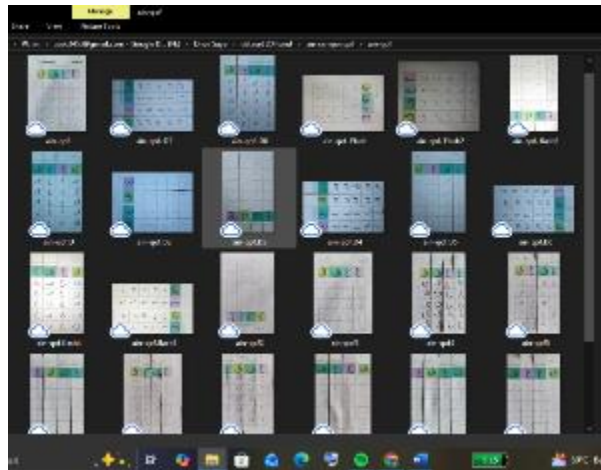
- TP (*True Positive*) adalah jumlah objek yang terdeteksi dengan benar.
- FP (*False Positive*) adalah jumlah objek yang terdeteksi tetapi sebenarnya bukan objek yang dimaksud.
- FN (*False Negative*) adalah objek yang seharusnya terdeteksi tetapi tidak berhasil di deteksi oleh model.

Selain itu, performa deteksi objek secara keseluruhan diukur menggunakan *mean Average Precision (mAP)* berdasarkan kurva *precision*, *recall* pada setiap objek. Selain metrik akurasi, penelitian ini juga mengukur kecepatan inferensi model menggunakan indikator *Frames per Second (FPS)* untuk mengetahui kemampuan sistem dalam melakukan seteksi huruf secara *real-time* [19]. Analisis performa model pada setiap kelas huruf hijaiyah dilakukan menggunakan *confusion matrix* yang menunjukkan hubungan antara kelas prediksi dan label sebenarnya. Penggunaan *mAP@0.5* digunakan untuk mengukur kemampuan umum model dalam mendeteksi objek, sedangkan *mAP@0.5:0.95* digunakan untuk mengevaluasi ketepatan lokalisasi *bounding box* pada berbagai tingkat *Intersection over Union (IoU)* yang lebih ketat.

### 3. HASIL DAN PEMBAHASAN

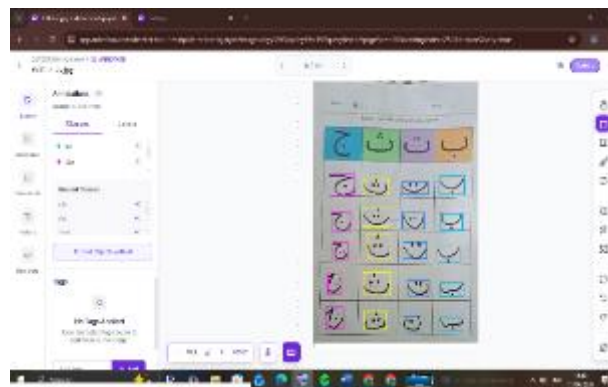
#### 3.1 Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 429 citra tulisan tangan huruf hijaiyah yang dikumpulkan dari siswa MDT Ula Addiya Kuwayuhan dengan rentang usia 4 sampai 6 tahun. Meskipun jumlah citra relatif terbatas, setiap citra mengandung beberapa objek huruf sehingga total objek yang dianotasi mencapai lebih dari 3.250 objek dalam 20 kelas. Kondisi ini menunjukkan bahwa dataset memiliki karakteristik *multi-object* dengan kepadatan objek yang tinggi dalam setiap citra. Contoh dataset mentah dapat dilihat pada Gambar 2.



**Gambar 2.** Dataset Mentah

Berdasarkan **Gambar 2**, setiap citra memuat lebih dari satu objek huruf sehingga dataset berada pada skenario *multi-object detection*. Tulisan tangan siswa menunjukkan variasi yang signifikan dalam bentuk, ukuran, dan ketebalan garis huruf, sehingga dataset memiliki karakteristik yang beragam [4]. Variasi bentuk tulisan, ukuran huruf, serta karakteristik multi-objekt pada dataset meningkatkan kompleksitas data yang digunakan dalam proses pelatihan model. Sebelum digunakan pada tahap pelatihan, dataset dipersiapkan melalui proses anotasi. Setiap objek huruf pada citra diberi label menggunakan teknik *bounding box* untuk merepresentasikan lokasi dan kelas objek secara terstruktur. Proses anotasi dilakukan menggunakan *platform* Roboflow. Contoh hasil anotasi dapat dilihat pada Gambar 3.



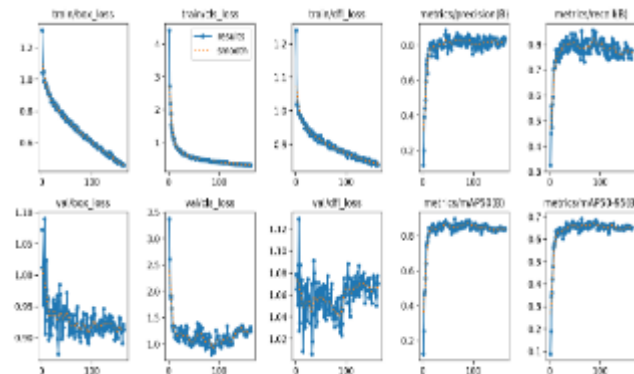
**Gambar 3.** Hasil Anotasi Data

Berdasarkan Gambar 3, setiap objek huruf telah diberi anotasi berupa *bounding box* yang merepresentasikan lokasi dan kelas huruf dalam citra. Namun, dalam beberapa kasus, jarak antar huruf yang berdekatan serta variasi ukuran tulisan menyebabkan terjadinya tumpang tindih (*overlapping*) antar *bounding box*. Kondisi *overlapping* antar objek meningkatkan kompleksitas dataset, khususnya pada proses lokalisasi objek.

Selain sebagai tantangan, keberagaman dan kompleksitas dataset ini juga berperan dalam menguji kemampuan generalisasi model terhadap variasi data di dunia nyata. Namun demikian, jumlah citra yang relatif terbatas tetap menjadi salah satu keterbatasan dalam penelitian ini, karena berpotensi memengaruhi kemampuan model dalam melakukan generalisasi terhadap data dengan variasi yang lebih luas.

### 3.2 Pelatihan Model

Pelatihan model dilakukan menggunakan platform *Google Colab* dengan dukungan GPU Tesla T4 untuk mempercepat proses komputasi. Selama proses pelatihan, performa model dipantau menggunakan metrik evaluasi yang umum digunakan dalam deteksi objek, yaitu *precision*, *recall*, serta *mean Average Precision (mAP)*[16]. Hasil pelatihan model YOLOv11s ditunjukkan pada Gambar 4.



**Gambar 4.** Gambar Training YOLO

Berdasarkan Gambar 4, terlihat bahwa nilai training loss yang terdiri dari *box loss*, *classification loss*, dan *distribution focal loss (DFL)* mengalami penurunan secara signifikan pada fase awal pelatihan dan kemudian cenderung stabil seiring bertambahnya *epoch*. Hal ini menunjukkan bahwa model mampu mempelajari representasi fitur visual huruf hijaiyah secara efektif. Pada sisi validasi, nilai *validation loss* juga menunjukkan tren penurunan meskipun dengan *fluktuasi* yang lebih tinggi, yang mengindikasikan adanya variasi data yang cukup kompleks, terutama akibat karakteristik tulisan tangan anak yang tidak seragam.

Selain itu, kurva *precision* dan *recall* menunjukkan peningkatan yang cepat pada awal pelatihan dan kemudian mencapai kondisi relatif stabil. Hal ini menandakan bahwa model secara bertahap mampu meningkatkan kemampuan dalam mendeteksi dan mengklasifikasikan objek huruf. Berdasarkan hasil evaluasi akhir, model memperoleh nilai *precision* sebesar 0,867 dan *recall* sebesar 0,831, yang menunjukkan keseimbangan performa yang cukup baik antara ketepatan deteksi dan kemampuan menemukan objek [16].

Nilai  $mAP@0.5$  sebesar 0,903 menunjukkan bahwa model memiliki kemampuan deteksi objek yang baik dalam mengenali huruf hijaiyah. Namun demikian, nilai  $mAP@0.5:0.95$  yang lebih rendah, yaitu sebesar 0,700, mengindikasikan bahwa performa model masih mengalami penurunan pada ambang *Intersection over Union (IoU)* yang lebih ketat. Hal ini menunjukkan bahwa meskipun model mampu mendeteksi keberadaan objek dengan baik, ketepatan dalam menentukan lokasi *bounding box* secara presisi masih perlu ditingkatkan[18]. Hasil tersebut menunjukkan bahwa model mampu memperthankan performa yang cukup baik meskipun dihadapkan dengan dataset dengan karakteristik multi-object dan variasi tulisan tangan yang tinggi.

Perbedaan nilai *precision* dan *recall* menunjukkan bahwa model cenderung lebih konservatif dalam melakukan prediksi, sehingga tingkat *false positive* yang relatif lebih rendah dibandingkan *false negative*. Kondisi ini menunjukkan bahwa model masih berpotensi melewatkan beberapa objek pada tulisan yang kurang jelas atau Dari sisi efisiensi, model menunjukkan kecepatan inferensi sebesar 12,3 ms per citra atau sekitar 80 *Frames Per Second (FPS)*, yang menunjukkan bahwa sistem mampu melakukan deteksi secara *real-time*. Kecepatan ini menjadi salah satu keunggulan utama pendekatan YOLO dalam implementasi sistem berbasis interaktif[18].

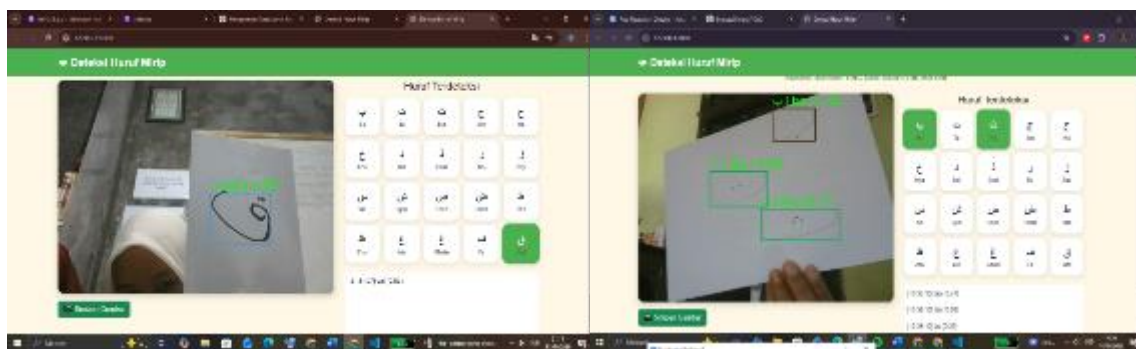


dapat ditingkatkan, yang konsisten dengan hasil evaluasi sebelumnya, khususnya pada nilai  $mAP@0.5:0.95$  yang lebih rendah dibandingkan  $mAP@0.5$  [19].

Untuk meningkatkan performa model, penelitian selanjutnya dapat di fokuskan pada penambahan jumlah dan keragaman dataset, khususnya pada huruf dengan kemiripan visual tinggi, serta perbaikan kualitas anotasi untuk meningkatkan presisi lokalisasi objek.

### 3.4 Implementasi Sistem

Model YOLOv11s yang telah dilatih kemudian diimplementasikan ke dalam sistem deteksi huruf hijaiyah berbasis web menggunakan *framework Flask*. *Framework Flask* dipilih karena bersifat ringan, fleksibel, serta mampu mengintegrasikan model *machine learning* ke dalam aplikasi *web* secara efisien [23]. Sistem yang dikembangkan mampu melakukan deteksi secara real-time dengan menampilkan *bounding box*, label kelas, serta nilai *confidence* pada setiap objek yang terdeteksi. Hasil implementasi sistem ditunjukkan pada Gambar 6 dan Gambar 7.



Gambar 6. Hasil Prediksi Benar

Gambar 7. Hasil Prediksi Salah

Pada **Gambar 6** ditunjukkan contoh hasil deteksi yang benar, di mana model mampu mengenali huruf dengan baik pada kondisi nyata (*in-the-wild*). Hal ini menunjukkan bahwa model telah mampu menangkap pola visual huruf hijaiyah meskipun terdapat variasi pada tulisan tangan yang digunakan. Namun demikian, model belum sepenuhnya konsisten dalam melakukan deteksi. Hal ini terlihat pada **Gambar 7** yang menunjukkan contoh kesalahan prediksi, terutama pada huruf yang memiliki kemiripan bentuk visual tinggi, seperti *'ain* (ع) dan *ghain* (غ), serta *ba* (ب), *ta* (ت), dan *tsha* (ث) [24]. Fenomena ini menunjukkan adanya keterbatasan model dalam membedakan fitur diskriminatif yang bersifat halus (*fine-grained features*), khususnya pada elemen kecil seperti titik.

Selain itu, dari perspektif deteksi objek, kesalahan prediksi juga dipengaruhi oleh ketidaktepatan lokalisasi *bounding box*. Hal ini konsisten dengan hasil evaluasi sebelumnya, di mana nilai  $mAP@0.5:0.95$  lebih rendah dibandingkan  $mAP@0.5$ , yang menunjukkan bahwa performa model menurun pada ambang *Intersection over Union* (IoU) yang lebih ketat [19]. Keterbatasan lain juga berasal dari jumlah dan keragaman dataset yang masih terbatas sehingga variasi tulisan tangan dunia nyata belum sepenuhnya terwakili. Selain itu, faktor eksternal seperti pencahayaan dan kompleksitas latar belakang citra juga dapat memengaruhi kualitas prediksi, terutama dalam menurunkan nilai *confidence* dan stabilitas deteksi.

Secara keseluruhan, implementasi sistem menunjukkan bahwa model YOLOv11s dapat diterapkan secara efektif dalam aplikasi berbasis web dengan kemampuan deteksi *real-time*. Sistem mampu memberikan umpan balik visual secara langsung, yang berpotensi mendukung proses pembelajaran interaktif. Meskipun demikian, masih terdapat keterbatasan dalam membedakan huruf dengan kemiripan visual tinggi, sehingga diperlukan pengembangan lebih lanjut, khususnya pada peningkatan kualitas dataset dan optimasi model, untuk meningkatkan akurasi dan konsistensi deteksi.

### 3.5 Perbandingan dengan Penelitian Sebelumnya

Penelitian ini memiliki beberapa perbedaan dibandingkan dengan penelitian sebelumnya dalam bidang pengenalan huruf Arab dan tulisan tangan berbasis kecerdasan buatan. Perbandingan penelitian ditunjukkan pada Tabel 3.

**Tabel 3.** Perbandingan Penelitian dengan Penelitian Sebelumnya

Penelitian	Metode	Fokus Penelitian	Jenis Deteksi
[13]	YOLOv9	Huruf Arab umum	Belum focus pada tulisan tangan anak usia dini
[14]	CNN	Klsifikasi huruf Arab	Hanya mendeteksi satu objek dalam satu citra
[15]	YOLO+TTS	Deteksi huruf alfabet dengan TTS	Belum diterapkan pada huruf hijaiyah
Penelitian ini	YOLOv11s	Deteksi multi-objek huruf hijaiyah tulisan tangan anak usia dini secara real-time	Masih terbatas pada jumlah dan variasi dataset

Berdasarkan **Tabel 3**. Penelitian sebelumnya umumnya masih berfokus pada klasifikasi satu karakter, penggunaan dataset yang lebih terstruktur, atau implementasi pada objek selain huruf hijaiyah tulisan tangan anak usia dini. Sementara itu penelitian ini difokuskan pada scenario multi-object handwritten hijaiyah dengan Tingkat kompleksitas yang lebih tinggi akibat variasi bentuk tulisan, ukuran huruf serta kemiripan visual antar kelas.

Selain itu, penelitian ini tidak hanya berfokus pada pengembangan model deteksi objek, tetapi juga mengimplementasikan sistem berbasis *web* secara *real-time* yang dilengkapi fitur *Text-to-Speech (TTS)* sebagai media pembelajaran interaktif. Pendekatan ini memungkinkan sistem memberikan umpan balik visual dan audio secara langsung pada proses pembelajaran huruf hijaiyah.

Meskipun penelitian ini tidak melakukan perbandingan eksperimental secara langsung dengan model *object detection* lain, hasil evaluasi menunjukkan bahwa YOLOv11s mampu memberikan performa yang baik pada dataset dengan karakteristik multi-object dan variasi tulisan tangan yang tinggi. Nilai  $mAP@0.5$  sebesar 0,903 menunjukkan bahwa model mampu mendeteksi objek huruf dengan baik meskipun dihadapkan pada kondisi tulisan tangan yang tidak terstruktur. Dengan demikian, penelitian ini menunjukkan bahwa pendekatan YOLOv11s memiliki potensi yang baik untuk diterapkan pada skenario deteksi huruf hijaiyah tulisan tangan anak usia dini secara *real-time*.

## 4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan YOLOv11s untuk mendeteksi multi-objek huruf hijaiyah tulisan tangan anak usia dini secara *real-time* pada aplikasi berbasis *web*. Model mampu mengenali beberapa huruf dalam satu citra dengan nilai *precision* 0,867, *recall* 0,831,  $mAP@0.5$  sebesar 0,903, dan kecepatan inferensi sekitar 80 FPS. Hasil tersebut menunjukkan bahwa YOLOv11s cukup mampu menangani deteksi tulisan tangan huruf hijaiyah yang memiliki variasi visual tinggi dan pola tulisan yang tidak seragam. Kontribusi penelitian ini terletak pada penerapan *object detection* berbasis YOLOv11s pada tulisan tangan anak usia dini dengan Tingkat kompleksitas yang tinggi, terutama karena kemiripan bentuk antarhuruf, variasi ukuran tulisan, dan keberadaan beberapa objek dalam satu citra. Selain itu, sistem diimplementasikan dalam aplikasi *web* yang dilengkapi fitur *Text-To-Speech (TTS)* sebagai media pembelajaran interaktif. Namun keterbatasan penelitian ini pada jumlah dan variasi dataset yang digunakan sehingga kesalahan deteksi pada huruf dengan bentuk serupa masih ditemukan. Penelitian

selanjutnya dapat difokuskan pada penambahan dataset, peningkatan kualitas anotasi, dan optimasi model untuk meningkatkan akurasi serta kemampuan generalisasi model.

#### DAFTAR PUSTAKA

- [1] I. G. A. A. M. Swari and Didith Pramuditya Ambara, “Video Animasi Mengenal Huruf dan Angka untuk Menstimulus Kemampuan Kognitif dan Bahasa Anak Usia Dini,” *J. Pendidik. Anak Usia Dini Undiksha*, vol. 10, no. 1, pp. 163–172, 2022, doi: 10.23887/paud.v10i1.47346.
- [2] E. Sopiati and M. Amin, “Penerapan Metode Pembelajaran Visual untuk Meningkatkan Hasil Belajar Materi Huruf Hijaiyah Siswa Sekolah Dasar,” pp. 252–265, 2025.
- [3] J. Beno, A. . Silen, and M. Yanti, “Mengenalkan Huruf Hijaiyah Pada Anak Usia Dini,” *Braz Dent J.*, vol. 33, no. 1, pp. 1–12, 2022.
- [4] A. Gea and Refni Fajar Wati Zega, “Metode Pembelajaran Kreatif dalam Pendidikan Anak Usia Dini,” *Khirani J. Pendidik. Anak Usia Dini*, vol. 3, no. 1, pp. 209–219, 2025, doi: 10.47861/khirani.v3i1.1622.
- [5] K. O. Nur Amanah, A. Kurnia, and N. Nurdiansah, “Hubungan Aktivitas Integrasi Visual Motorik Dengan Handwriting Skills Anak Usia Dini,” *J. Early Child. Educ. Stud.*, vol. 4, no. 1, pp. 132–165, 2024, doi: 10.54180/joeces.2024.4.1.163-196.
- [6] Pitri Juwita, Irma Suryani Siregar, and Sartika Dewi Harahap, “Pengembangan Alat Edukatif Balok Iqro’ untuk Mengenalkan Huruf Hijaiyah pada Anak Usia Dini,” *Khirani J. Pendidik. Anak Usia Dini*, vol. 3, no. 1, pp. 109–120, 2025, doi: 10.47861/khirani.v3i1.1563.
- [7] S. Samidin and A. Fadjeri, “Klasifikasi Gambar Batu-Kertas-Gunting Menggunakan Convolutional Neural Network dengan Fungsi Callback untuk Mencegah Overfitting,” *J. Penelit. Inov.*, vol. 4, no. 2, pp. 785–794, Jun. 2024, doi: 10.54082/jupin.413.
- [8] A. Fadjeri, “Identifikasi Teks dari Citra Menggunakan Optical Character Recognition,” *J. Ilm. SINUS*, vol. 22, no. 2, p. 13, 2024, doi: 10.30646/sinus.v22i2.819.
- [9] R. Munfarid and A. Fadjeri, “Deteksi Pengenalan Rambu Lalu-lintas Berdasarkan Citra Menggunakan Algoritma Yolo di Daerah Istimewa Yogyakarta,” vol. 06, pp. 1–12, 2026.
- [10] M. Huang and W. Jiang, “DMS-YOLO: Small target detection algorithm based on YOLOv11,” 2026, doi: 10.1371/journal.pone.0341991.
- [11] X. Zhang and J. Juang, “UAV Real-Time Image Recognition Using Lightweight,” 2026.
- [12] S. F. Mauliansyah, R. Rendi, R. Rosita, M. Iqbal, and H. Assulamy, “Analisis Kebutuhan Media Pembelajaran Interaktif Berbasis Web,” *J. Locus Penelit. dan Pengabd.*, vol. 2, no. 3, pp. 224–236, 2023.
- [13] M. Ilham, Z. Firdaus, S. Saidah, and N. Ibrahim, “Klasifikasi Tulisan Tangan Bahasa Arab Menggunakan Teknik Yolo,” vol. 12, no. 4, pp. 5122–5124, 2025.
- [14] M. Alheraki, R. Al-Matham, and H. Al-Khalifa, “Handwritten Arabic Character Recognition for Children Writing Using CNN and Stroke Identification,” *Human-Centric Intell. Syst.*, vol. 3, no. 2, pp. 147–159, 2023.
- [15] A. R. Putri, R. Dewi, J. T. Elektro, and P. N. Padang, “Penerapan Metode Yolov5 dan Text-To- Speech untuk Aplikasi Pengenalan Abjad dan Objek Sekitar Pada Anak Usia Dini,” vol. 15, pp. 94–101, 2023.
- [16] R. Padilla, W. Passos, T. Dias, S. Netto, and E. da Silva, “A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit,” *Electronics*, vol. 10, no. 3, p. 279, 2020, doi: 10.3390/electronics10030279.
- [17] H. P. Tulili, A. Septiarini, and Hamdani, “Klasifikasi Penyakit Tanaman Daun Padi Menggunakan Metode Deep Learning dengan Teknik Transfer Learning MobileNet,” *J. Inform. dan Rekayasa Elektron.*, vol. 8, no. 2, 2025.
- [18] N. Sultana, M. A. I. Shuvo, and M. S. Islam, “Performance Evaluation Metrics for Object

- Detection Models,” *IEEE Access*, 2022, [Online]. Available: <file:///mnt/data/423-Article-Text-1575-1-10-20240315.pdf>
- [19] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv*, 2020, [Online]. Available: <https://arxiv.org/abs/2004.10934>
- [20] 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,” *arXiv*, 2022, [Online]. Available: <https://arxiv.org/abs/2207.02696>
- [21] R. Austin and A. Syahrian, “Analisis Performa Algoritma Supervised Learning terhadap Data Deskripsi dengan Representasi dan Parameter Tuning,” *J. Inform. dan Rekayasa Elektron.*, vol. 8, no. 2, 2025.
- [22] M. Alheraki, R. Al-Matham, and H. Al-Khalifa, “Handwritten Arabic Character Recognition for Children Writing Using Convolutional Neural Network and Stroke Identification,” *Human-Centric Intell. Syst.*, vol. 3, no. 2, pp. 147–159, 2023, doi: 10.1007/s44230-023-00024-4.
- [23] K. R. Nur Manab, E. P. Mandyartha, and A. M. Rizki, “Rancang Bangun Sistem Deteksi Huruf Rusia Berbasis Web Flask,” *Pros. Semin. Nas. Inform. Bela Negara*, vol. 2, pp. 156–160, 2021, doi: 10.33005/santika.v2i0.108.
- [24] A. T. Sahlol, “Arabic Handwritten Characters Recognition Using Convolutional Neural Networks,” *Appl. Sci.*, 2021.