

Pengenalan Pola Huruf Bahasa Isyarat Menggunakan *Framework You Only Look Once (YOLO)*

Tri Ahmad Jabar*¹, Rudi Heriansyah², Evi Purnamasari³

^{1,2,3}Teknik Informatika, Ilmu Komputer dan Sains, Universitas Indo Global Mandiri

¹2021110122@students.uigm.ac.id*, ²rudi@uigm.ac.id, ³evi.ps@uigm.ac.id

Abstract

Sign language is an essential form of visual communication for individuals with hearing and speech impairments. However, many people still lack an understanding of the Indonesian Sign Language System (SIBI), which creates communication barriers. This study aims to develop a recognition system for SIBI sign language letters using the You Only Look Once (YOLO) framework. Image data of hand gestures were collected from the author, annotated using Roboflow, and trained with the YOLOv11 algorithm. The detected letters were not only recognized individually but also arranged into sentences in real-time through camera input using sequential letter processing. The best model achieved a precision of 0.835, recall of 0.928, and mean Average Precision (mAP) with mAP@50 (IoU 50%) of 0.968 and mAP@50-95 (average across multiple IoU thresholds) of 0.774. The system also reached an average accuracy of 0.831 and an F1-score of 0.865 in recognizing letters. In real-time testing, the system successfully composed the sentence "VINA SEDANG MAKAN" with an accuracy of 86.6%. These results demonstrate that the system is capable not only of detecting individual letters but also of composing meaningful sentences. This study is expected to provide a significant contribution to the development of inclusive technology that bridges communication between individuals with hearing impairments and the general public, with potential applications in education, public services, and daily interactions

Keywords: sign language, SIBI, YOLO, letter detection, image recognition

Abstrak

Bahasa isyarat merupakan bentuk komunikasi visual yang penting bagi penyandang disabilitas rungu wicara. Namun, masih banyak masyarakat yang belum memahami Sistem Isyarat Bahasa Indonesia (SIBI), sehingga menimbulkan hambatan komunikasi. Penelitian ini bertujuan untuk mengembangkan sistem pengenalan huruf bahasa isyarat SIBI menggunakan framework *You Only Look Once (YOLO)*. Data citra huruf dikumpulkan dari tangan penulis, dianotasi menggunakan *Roboflow*, dan dilatih dengan algoritma YOLOv11. Hasil deteksi huruf tidak hanya dikenali secara individu, tetapi juga disusun menjadi kalimat secara *real-time* melalui input kamera menggunakan pemrosesan sekuens huruf. Model terbaik menunjukkan nilai *precision* sebesar 0,835, *recall* 0,928, serta *mean Average Precision (mAP)* dengan *mAP@50 (IoU 50%)* sebesar 0,968 dan *mAP@50-95* (rata-rata pada berbagai ambang IoU) sebesar 0,774. Sistem juga mencapai akurasi rata-rata 0,831 dan *F1-score* sebesar 0,865 dalam pengenalan huruf. Pada pengujian *real-time*, sistem berhasil menyusun kalimat sederhana "VINA SEDANG MAKAN" dengan akurasi 86,6%. Hasil ini membuktikan bahwa sistem tidak hanya mampu mendeteksi huruf, tetapi juga dapat merangkai huruf menjadi kalimat bermakna. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan teknologi inklusif yang dapat menjembatani komunikasi antara penyandang disabilitas rungu wicara dengan masyarakat umum, serta berpotensi diimplementasikan dalam bidang pendidikan, pelayanan publik, maupun aplikasi sehari-hari.

Kata kunci: bahasa isyarat, SIBI, YOLO, deteksi objek, pengolahan citra

©This work is licensed under a Creative Commons Attribution -ShareAlike 4.0 International License

1. Pendahuluan

Komunikasi merupakan aspek fundamental dalam kehidupan manusia, baik dalam konteks sosial, pendidikan, pekerjaan, maupun layanan publik. Setiap individu memiliki kebutuhan untuk menyampaikan ide, keinginan, dan informasi secara efektif. Tanpa komunikasi yang baik, banyak aspek kehidupan dapat terhambat, termasuk pembelajaran, pelayanan, serta partisipasi aktif dalam masyarakat [1]. Salah satu bentuk komunikasi yang digunakan oleh penyandang disabilitas rungu wicara adalah bahasa isyarat, bahasa ini menjadi sarana utama bagi mereka dalam menjalin interaksi sehari-hari [2]. Di Indonesia, bahasa isyarat yang digunakan secara resmi adalah Sistem Isyarat Bahasa Indonesia (SIBI), yang disusun berdasarkan

struktur gramatikal bahasa Indonesia [3]. SIBI membantu penyandang tunarungu untuk menyampaikan pesan dan memahami informasi dengan cara yang lebih sistematis.

Namun, dalam kenyataannya, masih banyak masyarakat yang belum memahami bahasa isyarat, khususnya SIBI. Kurangnya pemahaman ini menimbulkan hambatan komunikasi dua arah yang berdampak luas, misalnya kesulitan penyandang tunarungu dalam mengakses layanan kesehatan, pendidikan, maupun interaksi sosial sehari-hari [4]. Hambatan ini tidak hanya mempersempit ruang komunikasi, tetapi juga berpotensi menimbulkan kesenjangan sosial dan diskriminasi terhadap penyandang disabilitas rungu wicara. Menurut data dari

Badan Pusat Statistik (2020), terdapat lebih dari 22 juta penyandang disabilitas di Indonesia, dan sekitar 0,4% di antaranya merupakan penyandang disabilitas pendengaran. Angka ini menunjukkan bahwa terdapat ratusan ribu orang yang berpotensi mengalami hambatan komunikasi karena keterbatasan pemahaman terhadap SIBI.

Perkembangan teknologi dalam bidang *computer vision* dan *deep learning* membuka peluang untuk mengatasi hambatan komunikasi tersebut melalui pengembangan sistem pengenalan bahasa isyarat secara otomatis [5]. Teknologi ini memungkinkan terjemahan bahasa isyarat ke dalam teks atau suara secara cepat, sehingga dapat menjembatani komunikasi antara penyandang disabilitas rungu wicara dan masyarakat umum. Salah satu pendekatan yang banyak digunakan adalah algoritma *You Only Look Once* (YOLO), yaitu metode deteksi objek berbasis *deep learning* yang dikenal karena kecepatannya dan akurasi [6]. YOLO bekerja dengan memproses seluruh gambar dalam satu kali perhitungan, sehingga sangat cocok digunakan dalam aplikasi *real-time*. [6]. Dengan memanfaatkan arsitektur *convolutional neural network* (CNN), YOLO dapat melakukan deteksi lokasi dan klasifikasi objek secara bersamaan. Dalam konteks penelitian ini, YOLO digunakan untuk mengenali huruf-huruf bahasa isyarat SIBI dari citra tangan pengguna, kemudian menyusunnya menjadi kalimat teks secara otomatis.

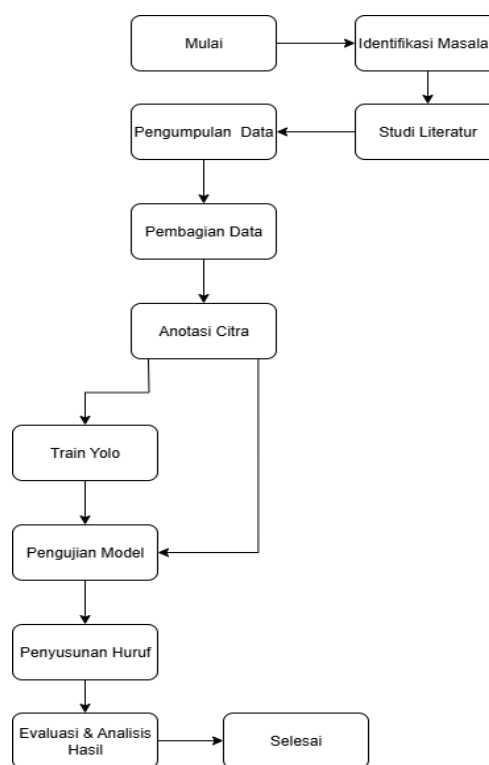
Sejumlah penelitian sebelumnya telah membahas pengenalan bahasa isyarat. Sari dan Nugroho (2022) mengembangkan sistem deteksi huruf SIBI menggunakan CNN, dengan akurasi sebesar 92% [7]. Namun, penelitian tersebut terbatas pada pengenalan huruf secara terpisah, belum menyusun huruf menjadi kata atau kalimat. Selanjutnya, penelitian oleh Imam dkk. (2023) menerapkan YOLOv5 untuk mendeteksi alfabet bahasa isyarat, dengan performa cukup baik pada kondisi cahaya rendah dan latar belakang kompleks, tetapi belum mengimplementasikan penyusunan huruf menjadi kalimat [8]. Penelitian lain oleh Nugroho dkk. (2023) mengombinasikan CNN dan RNN untuk mendeteksi bahasa isyarat berbasis video, namun hasilnya masih terbatas pada pengenalan kata tunggal. [9]. Penelitian lain, seperti Pratama dkk. (2024), telah menunjukkan efektivitas YOLO dalam mendeteksi bahasa isyarat SIBI secara *real-time*, tetapi implementasinya masih fokus pada deteksi objek huruf, bukan penyusunan kalimat [10].

Dari tinjauan tersebut dapat disimpulkan bahwa masih terdapat celah penelitian, yaitu bagaimana mengembangkan sistem pengenalan huruf bahasa isyarat yang tidak hanya mendeteksi huruf secara akurat, tetapi juga mampu menyusunnya menjadi kalimat teks secara *real-time*. Meskipun telah banyak penelitian yang menggunakan CNN dan YOLO dalam mengenali bahasa isyarat, sebagian besar dari penelitian tersebut masih terbatas pada pengolahan satu huruf atau kata secara terpisah. Sistem-sistem tersebut

belum sepenuhnya mampu menyusun rangkaian huruf menjadi kalimat utuh sebagaimana digunakan dalam komunikasi sehari-hari. Dalam praktik nyata, penyandang disabilitas tidak hanya menyampaikan satu huruf saja, tetapi menyusun huruf-huruf menjadi kata dan kalimat untuk menyampaikan pesan yang bermakna secara penuh [10]. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem pengenalan huruf bahasa isyarat SIBI berbasis YOLOv11 yang dapat digunakan secara praktis oleh masyarakat umum, serta mendukung komunikasi inklusif bagi penyandang disabilitas rungu wicara. Hal ini diharapkan dapat meningkatkan efektivitas komunikasi, terutama dalam konteks penggunaan sistem secara *real-time*, baik untuk keperluan pembelajaran, pelayanan publik, maupun komunikasi sosial.

2. Metode Penelitian

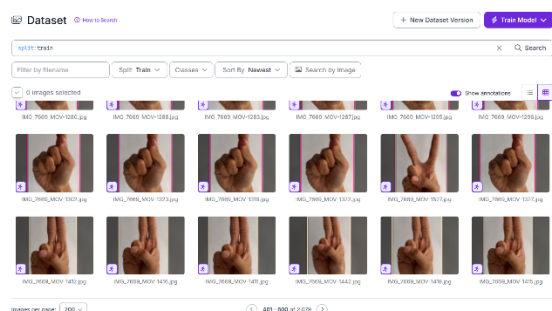
Penelitian ini termasuk dalam kategori penelitian pengembangan sistem (*research and development*) yang memanfaatkan pendekatan eksperimen untuk menguji kinerja model deteksi huruf bahasa isyarat berbasis *deep learning*. Desain penelitian terdiri dari beberapa tahapan utama, yaitu identifikasi masalah, pengumpulan data, pembagian data, anotasi data, pelatihan model, pengujian model, penyusunan huruf, evaluasi hasil, dan penyelesaian. Alirannya dapat dilihat pada Gambar 1.



Gambar 1. Database Mirroring Architecture

2.1. Pengumpulan Data

Pengumpulan data merupakan tahapan penting dalam penyediaan dataset untuk pelatihan dan pengujian model deteksi huruf bahasa isyarat. Data yang digunakan berupa gambar tangan yang membentuk huruf-huruf dalam Sistem Isyarat Bahasa Indonesia (SIBI), khususnya alfabet A hingga Z yang disampaikan menggunakan satu tangan [11]. Seluruh dataset yang digunakan merupakan data primer yang dikumpulkan secara langsung oleh peneliti menggunakan kamera smartphone dengan resolusi 1920×1080 piksel. Objek citra berupa tangan peneliti sendiri yang membentuk pose huruf-huruf isyarat SIBI. Untuk memastikan keberagaman dan representativitas data, pengambilan gambar dilakukan dalam berbagai variasi kondisi, antara lain pencahayaan alami di siang hari, pencahayaan redup dalam ruangan, dan pencahayaan buatan dengan lampu, serta dengan sudut pengambilan gambar dari depan lurus, selain itu, latar belakang juga divariasikan, mulai dari polos terang, polos gelap. Dengan variasi kondisi tersebut, dataset yang dihasilkan menjadi lebih bervariasi dan representatif, sehingga diharapkan dapat meningkatkan kemampuan model dalam mengenali huruf bahasa isyarat SIBI pada berbagai situasi nyata.



Gambar 2. Database Mirroring Architecture

2.2. Pembagian Data

Dataset yang telah dikumpulkan berjumlah Sebanyak 2860 gambar dan dataset dibagi ke dalam tiga kelompok, yaitu data latih sebanyak 2080 (73%) data gambar untuk proses pelatihan model, data validasi sebanyak 390 (14%) untuk memantau performa model selama pelatihan dan mencegah overfitting, serta data uji 390 (14%) data gambar untuk mengevaluasi akurasi model terhadap data baru yang belum pernah dilihat sebelumnya. Pembagian data secara proporsional ini bertujuan untuk membangun sistem yang optimal, akurat, dan mampu mengenali huruf bahasa isyarat secara umum [12]. Seluruh data yang dikumpulkan kemudian melalui tahap pra-pemrosesan dan anotasi menggunakan format standar YOLO sebelum digunakan dalam proses pelatihan model.



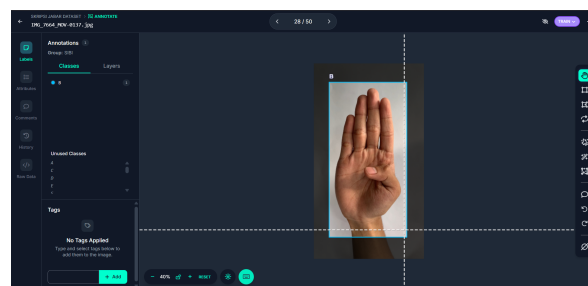
Gambar 3. Database Mirroring Architecture

2.3. Anotasi Citra

Pada tahap ini, proses pelabelan dilakukan untuk menandai posisi dan jenis huruf isyarat tangan yang muncul pada setiap gambar, serta mengelompokkannya ke dalam kelas yang sesuai. Dalam penelitian ini, terdapat 26 kelas yang digunakan, mewakili huruf A hingga Z dalam bahasa isyarat SIBI. Setiap gambar dianotasi dengan menggambar *bounding box* pada area tangan yang menunjukkan huruf tertentu, kemudian diberi label huruf sesuai dengan posisi tangan dalam citra.

Proses anotasi dilakukan menggunakan platform *Roboflow*, yang memudahkan peneliti dalam melakukan pelabelan secara manual dan sistematis [12]. Untuk memastikan ketepatan pelabelan, setiap citra diperiksa dan diverifikasi kembali secara berulang oleh peneliti, sehingga posisi *bounding box* dan label kelas benar-benar sesuai dengan bentuk huruf yang ditampilkan. Validasi ini penting dilakukan agar tidak terjadi kesalahan pelabelan yang dapat memengaruhi hasil pelatihan. [13].

Kualitas data berperan sangat penting dalam keberhasilan model. Kesalahan pelabelan, seperti posisi *bounding box* yang tidak tepat atau label huruf yang keliru, dapat menyebabkan model YOLO mempelajari pola yang salah sehingga menurunkan tingkat akurasi pada saat pengujian. Oleh karena itu, semua data yang telah dianotasi diperiksa ulang sebelum diekspor ke dalam format YOLOv11. Dalam format ini, setiap gambar disertai file .txt yang berisi informasi kelas objek, koordinat tengah objek (x, y), serta ukuran *bounding box* (width dan height) dalam skala relatif terhadap dimensi gambar [14]. Setelah semua data selesai dianotasi dan diverifikasi, dataset dinyatakan siap untuk digunakan dalam proses pelatihan model YOLO.



Gambar 4. Proses Labelling Gambar

2.4. Pelatihan Yolo

Pada tahap pelatihan, model YOLOv11 digunakan untuk melakukan proses deteksi objek berupa huruf-huruf bahasa isyarat SIBI. Dataset yang telah dianotasi sebelumnya dimasukkan ke dalam lingkungan pelatihan menggunakan *Roboflow* dan *Google Colaboratory*. Pelatihan dilakukan dengan konfigurasi parameter yang telah disesuaikan, termasuk resolusi citra sebesar 640×640 piksel, *batch size*, *learning rate*, dan jumlah *epoch*. Resolusi ini dipilih karena ukuran tersebut merupakan standar pada YOLO yang mampu

menyeimbangkan antara detail citra dan kecepatan komputasi. Parameter *batch size* dan *learning rate* menggunakan nilai default dari *framework Ultralytics* untuk menjaga stabilitas pelatihan. Jumlah *epoch* divariasikan hingga 140 kali untuk melihat pengaruh lamanya pelatihan terhadap performa model. Model dilatih menggunakan arsitektur YOLOv11 karena versi ini menawarkan kecepatan dan akurasi deteksi yang tinggi dalam satu proses (*single-stage detector*) [15]. Untuk meningkatkan kualitas pelatihan, digunakan teknik *early stopping*, yaitu menghentikan proses pelatihan ketika performa validasi tidak lagi menunjukkan peningkatan signifikan. Pendekatan ini bertujuan untuk mencegah terjadinya *overfitting* dan sekaligus mempercepat waktu pelatihan.

Proses pelatihan dilakukan dengan memanfaatkan GPU gratis yang disediakan oleh *Google Colab* untuk mempercepat waktu komputasi. Selama proses pelatihan, sistem secara otomatis mempelajari pola-pola visual dari huruf-huruf tangan berdasarkan bounding box yang diberikan [16]. Evaluasi dilakukan pada setiap *epoch* dengan metrik evaluasi seperti *precision*, *recall*, dan *mean Average Precision (mAP)* [17]. Visualisasi seperti *confusion matrix* juga digunakan untuk melihat sebaran performa antar huruf kelas.

2.5. Pengujian Model Yolo

Setelah model YOLOv11 dilatih, dilakukan tahap pengujian menggunakan data uji berupa citra tangan berbahasa isyarat SIBI yang berasal dari data primer, yaitu citra tangan milik penulis sendiri yang belum pernah digunakan dalam proses pelatihan. Tujuan dari tahap ini adalah untuk mengukur kemampuan generalisasi model dalam mengenali huruf secara akurat pada data baru yang bersifat nyata. Model akan memproses citra uji dan menghasilkan prediksi berupa *bounding box*, label huruf, serta nilai *confidence score* pada setiap huruf yang terdeteksi [18].

Evaluasi performa model dilakukan menggunakan beberapa metrik utama, yaitu *precision*, *recall*, *F1-score*, dan *mean Average Precision (mAP)*. *Precision* digunakan untuk mengukur sejauh mana prediksi huruf yang benar dibandingkan dengan seluruh prediksi yang dihasilkan model. *Recall* mengukur kemampuan model dalam mendeteksi semua huruf yang benar dari seluruh huruf yang ada pada citra uji. Kedua metrik ini saling melengkapi, karena *precision* yang tinggi menunjukkan minimnya kesalahan prediksi, sedangkan *recall* yang tinggi menunjukkan kemampuan model menangkap lebih banyak huruf yang relevan. *F1-score* kemudian digunakan sebagai ukuran keseimbangan antara *precision* dan *recall*, sehingga memberikan gambaran performa model secara lebih komprehensif.

Selain itu, mAP digunakan untuk menilai akurasi deteksi berdasarkan berbagai ambang batas *Intersection over Union (IoU)*. Nilai mAP@50 menggambarkan tingkat akurasi model pada ambang IoU 50%, sedangkan mAP@50–95 menunjukkan rata-

rata akurasi model pada berbagai ambang batas yang lebih ketat. Dengan kombinasi metrik tersebut, dapat diperoleh gambaran yang menyeluruh mengenai keunggulan maupun keterbatasan model dalam mendeteksi huruf bahasa isyarat SIBI secara real-time.

2.6. Penyusunan Huruf

Penyusunan huruf merupakan tahap penting untuk menilai keberhasilan sistem dalam mencapai tujuan utama, yaitu menerjemahkan huruf bahasa isyarat SIBI menjadi teks yang dapat dipahami masyarakat umum. Implementasi dilakukan menggunakan bahasa pemrograman *Python* dengan bantuan pustaka *OpenCV* untuk mengakses kamera secara real-time dan menampilkan hasil prediksi secara langsung. Setiap frame video yang masuk dari kamera diproses oleh model YOLOv11 untuk mendeteksi gesture huruf SIBI. Hasil prediksi berupa huruf dan nilai *confidence score* kemudian diproses lebih lanjut untuk membentuk rangkaian kata atau kalimat.

Agar sistem dapat bekerja dengan stabil dalam kondisi *real-time*, diterapkan mekanisme pengaturan interval deteksi, yaitu huruf hanya dicatat setiap 3 detik sekali. Pendekatan ini mencegah sistem mencatat huruf secara berulang akibat frame yang berdekatan. Selain itu, sistem juga membandingkan hasil prediksi huruf saat ini dengan huruf sebelumnya, sehingga huruf yang sama tidak direkam dua kali apabila pengguna masih berada pada gesture yang sama.

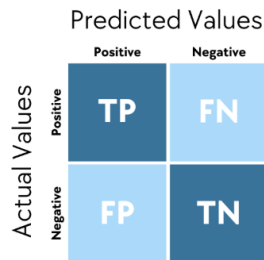
Untuk mengurangi kesalahan akibat gerakan tangan yang tidak sempurna atau tidak jelas, sistem memanfaatkan nilai *confidence score* yang dihasilkan model. Jika nilai prediksi berada di bawah ambang batas tertentu, maka huruf tersebut tidak dicatat agar tidak mengganggu keakuratan susunan kalimat. Semua huruf yang berhasil dikenali akan dikumpulkan dalam bentuk string hasil_kalimat, ditampilkan secara langsung di layar, dan disimpan ke dalam file teks sebagai hasil akhir. Sistem juga dilengkapi dengan fungsi interaktif sederhana, di mana pengguna dapat menekan tombol tertentu (misalnya r untuk mereset kalimat dan q untuk keluar dari program) sehingga proses penyusunan kalimat dapat lebih fleksibel dan mudah digunakan.

2.7. Evaluasi dan Analisis Hasil

Evaluasi sistem dilakukan untuk mengukur performa model YOLOv11 dalam mengenali huruf-huruf bahasa isyarat SIBI. Pengujian dilakukan menggunakan dataset uji yang berbeda dari data latih untuk memastikan kemampuan generalisasi model. Selama proses evaluasi, digunakan *confusion matrix* untuk melihat distribusi prediksi terhadap kelas yang sebenarnya. *Confusion matrix* membantu dalam mengidentifikasi huruf-huruf yang sering mengalami kesalahan klasifikasi dengan cara membandingkan hasil prediksi model dengan label sebenarnya.

Secara umum, confusion matrix terdiri dari empat komponen utama, yaitu *True Positive (TP)*, prediksi

benar pada kelas positif; *True Negative* (TN), prediksi benar pada kelas negatif; *False Positive* (FP), prediksi salah ketika model mengklasifikasikan data negatif sebagai positif; serta *False Negative* (FN), prediksi salah ketika data positif tidak dikenali oleh model. Dengan melihat sebaran nilai TP, FP, TN, dan FN pada setiap kelas huruf, peneliti dapat mengetahui huruf mana yang sering tertukar, misalnya huruf dengan bentuk *gesture* yang mirip. Informasi ini menjadi penting untuk menganalisis kelemahan model serta potensi perbaikan dataset maupun parameter pelatihan [16].



Gambar 5. Database Mirroring Architecture

Setelah perbandingan dilakukan *confusion matrix*, pengujian dilakukan menggunakan metrik *precision*, *recall*, *accuracy* dan *mean Average Precision (mAP)* pada dua cakupan, yaitu *mAP@50 (IoU 50%)* dan *mAP@50–95* menggambarkan tingkat keakuratan model dalam mengklasifikasikan data dengan benar [19].

Precision adalah metrik yang mengukur sejauh mana hasil prediksi positif dari model benar adanya [13]. Dengan kata lain, *precision* menunjukkan proporsi prediksi yang benar dari seluruh hasil prediksi positif yang dibuat oleh model.

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

Recall, yang juga dikenal sebagai sensitivitas atau *true positive rate*, mengukur sejauh mana model mampu mengenali semua data aktual yang termasuk ke dalam kelas positif [13]. *Recall* menunjukkan proporsi dari data positif yang berhasil dikenali secara benar oleh model.

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

Accuracy adalah metrik yang paling umum digunakan, dan menunjukkan proporsi prediksi yang benar (baik positif maupun negatif) terhadap seluruh prediksi yang dilakukan [13].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

3. Hasil dan Pembahasan

Bagian ini membahas hasil penelitian yang mencakup proses pelatihan model YOLOv11, evaluasi kinerja deteksi huruf bahasa isyarat SIBI, serta implementasi sistem dalam kondisi *real-time*. Tujuan bab ini adalah untuk menampilkan performa sistem yang dikembangkan sekaligus menganalisis keunggulan dan

keterbatasannya dalam menjawab permasalahan komunikasi penyandang disabilitas rungu wicara.

3.1. Pelatihan Menggunakan Yolo

Pelatihan dilakukan di lingkungan *Google Colab* dengan memanfaatkan GPU agar proses komputasi berjalan lebih efisien. Pada tahap ini, dilakukan eksperimen terhadap beberapa variasi jumlah epoch, yaitu 20, 40, 60, 80, 100, 120, dan 140 *epoch*. Pemilihan variasi tersebut dimaksudkan untuk mengevaluasi pengaruh durasi pelatihan terhadap kemampuan model dalam mengenali huruf-huruf bahasa isyarat SIBI. *Epoch* kecil (20–40) digunakan untuk melihat performa model pada tahap awal pelatihan, epoch menengah (60–100) digunakan untuk mengevaluasi stabilitas performa, sedangkan epoch besar (120–140) digunakan untuk menguji apakah model mengalami peningkatan signifikan atau justru *overfitting*. Dengan cara ini, dapat diperoleh gambaran mengenai titik optimal jumlah *epoch* bagi model.

Parameter lain disesuaikan dengan konfigurasi standar YOLOv11 untuk menjaga konsistensi pelatihan. Resolusi citra ditetapkan pada 640×640 piksel karena ukuran ini seimbang antara detail visual dan efisiensi komputasi. *Batch size* diatur agar sesuai dengan kapasitas GPU *Google Colab* sehingga tidak menimbulkan *out of memory error*, sedangkan learning rate mengikuti nilai default yang telah dioptimalkan oleh YOLOv11 untuk deteksi objek umum. Dalam penelitian ini tidak digunakan teknik augmentasi tambahan, tetapi untuk mencegah *overfitting* diterapkan mekanisme *early stopping*, sehingga pelatihan dapat berhenti lebih awal apabila performa model tidak menunjukkan peningkatan signifikan pada validasi. Hasil dari proses pelatihan mencakup metrik *precision*, *recall*, *accuracy*, dan *mean Average Precision (mAP@50 dan mAP@50–95)* yang akan digunakan untuk membandingkan performa antar variasi epoch serta menentukan konfigurasi model terbaik.

Tabel 1. Tabel Hasil Pelatihan

Epoch	Time	Precision	Recall	Accuracy	mAP@50–95
20	0.221	0.853	0.952	0.869	0.777
40	0.426	0.894	0.908	0.907	0.774
60	0.688	0.872	0.909	0.890	0.792
80	0.887	0.917	0.852	0.884	0.792
100	1.100	0.958	0.911	0.920	0.801
120	1.284	0.882	0.916	0.879	0.798
140	1.562	0.908	0.865	0.884	0.788

Berdasarkan hasil pada Tabel 1 memperlihatkan adanya fluktuasi nilai *precision*, *recall*, *accuracy*, dan *mAP* seiring dengan peningkatan jumlah *epoch*. Pada epoch awal (20–40), *recall* relatif tinggi (0.952 → 0.908) namun *precision* masih moderat (0.853 → 0.894). Hal ini menunjukkan bahwa pada tahap awal pelatihan model cenderung lebih inklusif dalam mendeteksi huruf (lebih banyak benar terdeteksi),

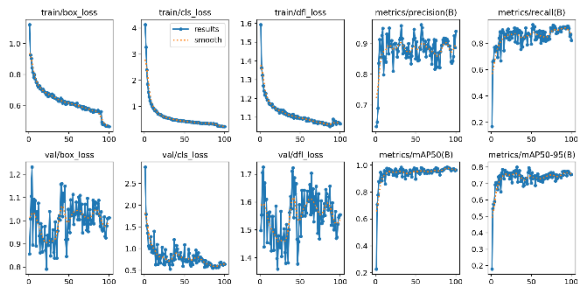
tetapi masih menghasilkan beberapa prediksi salah (*false positive*).

Pada *epoch* menengah (60–80), nilai *precision* dan *recall* mulai menunjukkan pergeseran. *Precision* meningkat (hingga 0.917), tetapi *recall* sedikit menurun (0.852–0.909). Kondisi ini menggambarkan bahwa model mulai lebih selektif dalam klasifikasi, meskipun berisiko melewatkan beberapa huruf yang benar (*false negative*). Nilai *mAP* stabil pada kisaran 0.792, menandakan konsistensi performa deteksi objek.

Performa optimal dicapai pada *epoch* 100, di mana *precision* tertinggi (0.958), *recall* relatif stabil (0.911), *accuracy* meningkat menjadi 0.920, dan *mAP@50–95* mencapai puncaknya (0.801). Hasil ini menunjukkan keseimbangan terbaik antara deteksi yang benar, ketepatan klasifikasi, dan kemampuan model mengenali posisi objek.

Namun, pada *epoch* 120–140, terlihat adanya penurunan performa (*precision* turun ke 0.882–0.908, *recall* 0.865–0.916, dan *mAP* menurun ke 0.788). Hal ini dapat mengindikasikan mulai terjadinya *overfitting*, di mana model terlalu menyesuaikan diri dengan data latih sehingga performanya pada data uji tidak lagi meningkat.

Dengan mempertimbangkan tren keseluruhan, *epoch* 100 dipilih sebagai model terbaik karena memberikan hasil paling seimbang dan stabil pada semua metrik evaluasi. Faktor utama yang mendukung capaian ini adalah kecukupan iterasi pelatihan untuk mengenali pola data, serta penerapan *early stopping* yang mencegah pelatihan berlebihan.



Gambar 6. Hasil Grafik Pelatihan epoch 100

3.2. Pengujian Menggunakan Yolo

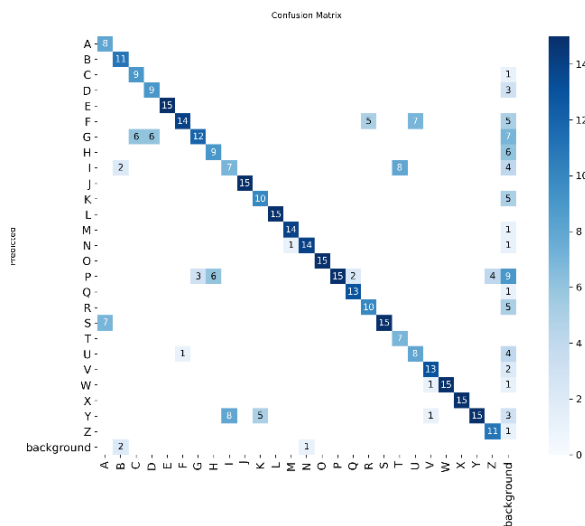
Setelah proses pelatihan selesai, dilakukan tahap pengujian untuk mengevaluasi sejauh mana model dapat mendeteksi huruf-huruf bahasa isyarat yang belum pernah dilihat sebelumnya. Pengujian dilakukan menggunakan model hasil pelatihan terbaik, yaitu model pada *epoch* ke-100, yang sebelumnya telah menunjukkan performa paling optimal berdasarkan metrik evaluasi pada data validasi. Model ini dipilih karena menghasilkan nilai *mAP@50–95* tertinggi selama proses pelatihan, yang menandakan kestabilan dan kemampuan generalisasi model yang baik. Selama proses pengujian, model menghasilkan sejumlah metrik performa yang penting untuk mengevaluasi kinerjanya, antara lain *precision*, *recall*, *mAP@50*, dan

mAP@50–95. Metrik-metrik tersebut menunjukkan seberapa tepat dan konsisten model dalam mendeteksi dan mengklasifikasikan huruf dari gambar input. Selain itu, pengujian juga menghasilkan *confusion matrix*, yang memberikan gambaran distribusi hasil klasifikasi untuk setiap kelas huruf, baik yang terdeteksi dengan benar maupun yang mengalami kesalahan prediksi, dan hasil Pengujian Model dapat dilihat pada Gambar 7.

Class	Images	Instances	Box(P)	R	mAP50	mAP50-95
all	390	390	0.835	0.928	0.968	0.774
A	15	15	1	0.591	0.995	0.845
B	15	15	1	0.855	0.987	0.894
C	15	15	1	0.89	0.995	0.736
D	15	15	1	0.89	0.995	0.846
E	15	15	0.948	1	0.995	0.906
F	15	15	0.453	1	0.948	0.893
G	15	15	0.474	1	0.767	0.505
H	15	15	0.721	0.8	0.922	0.529
I	15	15	0.521	0.8	0.758	0.681
J	15	15	0.944	1	0.995	0.847
K	15	15	0.977	1	0.995	0.902
L	15	15	0.908	1	0.995	0.881
M	15	15	0.967	1	0.995	0.776
N	15	15	0.857	1	0.988	0.841
O	15	15	0.953	1	0.995	0.752
P	15	15	0.369	1	0.995	0.552
Q	15	15	0.935	0.954	0.982	0.616
R	15	15	0.969	1	0.995	0.88
S	15	15	0.653	1	0.995	0.817
T	15	15	1	0.563	0.995	0.892
U	15	15	0.874	0.929	0.957	0.763
V	15	15	0.979	1	0.995	0.878
W	15	15	0.831	1	0.995	0.846
X	15	15	0.942	1	0.995	0.806
Y	15	15	0.449	1	0.952	0.782
Z	15	15	1	0.867	0.995	0.461

Gambar 7. Pengujian Performa Model

Selama proses pengujian, model menghasilkan sejumlah metrik performa, antara lain *precision* sebesar 0,835, *recall* 0,928, *accuracy* 0,831, *mAP@50* sebesar 0,968, dan *mAP@50–95* sebesar 0,774. Hasil ini mencerminkan keseimbangan antara presisi deteksi dan konsistensi model dalam mengenali huruf secara umum.



Gambar 8. Hasil Confusion Matrix

Selain itu, pengujian juga menghasilkan confusion matrix pada Gambar 8 yang memberikan gambaran distribusi klasifikasi untuk setiap huruf. Secara umum, diagonal utama pada confusion matrix menunjukkan dominasi prediksi yang benar, menandakan bahwa sebagian besar huruf dapat dikenali dengan baik. Namun, terdapat beberapa huruf yang masih rawan

mengalami kesalahan klasifikasi. Misalnya, huruf A sering tertukar dengan huruf S, serta huruf I dan T yang kerap saling tertukar. Kesalahan ini dapat disebabkan oleh kemiripan bentuk gesture tangan antar huruf, serta faktor pencahayaan dan latar belakang saat pengambilan data. Sebaliknya, huruf-huruf seperti J, K, L, W, X, Y, dan Z menunjukkan performa stabil dengan akurasi tinggi.

3.3. Penyusunan Huruf Secara *Realtime*

Tahap selanjutnya adalah melakukan pengujian sistem secara *real-time* untuk mengevaluasi kemampuan model dalam mendeteksi huruf bahasa isyarat SIBI dan menyusunnya menjadi kalimat utuh. Model yang digunakan adalah model hasil pelatihan pada epoch ke-100, yang sebelumnya dipilih sebagai model terbaik berdasarkan evaluasi metrik *precision*, *recall*, dan *mAP*.

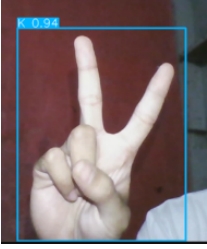
Implementasi sistem dilakukan menggunakan *library OpenCV* untuk menangkap citra dari kamera laptop secara berkelanjutan. Setiap frame video diproses oleh model YOLOv11 untuk mendeteksi gesture huruf. Agar hasil penyusunan kalimat tidak dipengaruhi oleh duplikasi huruf dari frame yang berdekatan, sistem menerapkan mekanisme jeda deteksi sekitar 3 detik. Selain itu, sistem hanya mencatat huruf baru jika berbeda dari huruf yang terdeteksi sebelumnya. Pendekatan ini memastikan bahwa penyusunan huruf menjadi kalimat dapat dilakukan secara lebih stabil dan menyerupai alur komunikasi nyata. Percobaan dilakukan dalam kondisi *real-time* menggunakan kamera laptop, di mana pengguna secara bergantian membentuk gestur huruf dari kalimat, yaitu:

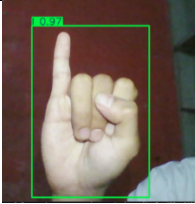



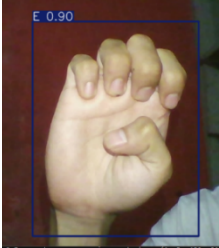


Tabel 2. Kalimat yang dipakai untuk Penyusunan Huruf



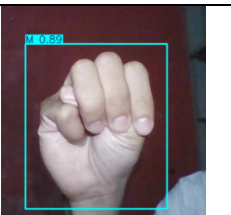


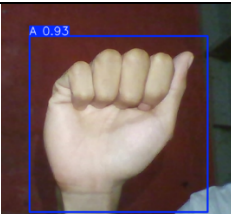

Target Kalimat
VINA SEDANG MAKAN

Masing-masing huruf dari kalimat divisualisasikan secara berurutan melalui gerakan tangan, dan sistem mendeteksi huruf secara otomatis berdasarkan model YOLO yang telah dilatih sebelumnya. Berikut hasil deteksi penyusunan huruf pada Tabel 3.

Tabel 3. Hasil Deteksi Kalimat

Huruf	Hasil Deteksi	Output	Keterangan
V		K	Salah

I		I	Benar
N		N	Benar
A		A	Benar
S		I	Salah
E		E	Benar
D		D	Benar
A		A	Benar

N		N	Benar
G		G	Benar
M		M	Benar
A		A	Benar
K		K	Benar
A		K	Benar
N		N	Benar

Speed: 1.4ms preprocess, 63.7ms inference
Kalimat akhir: KINA IEDANG MAKAN

Gambar 9. Hasil Akhir Penyusunan Kalimat

Berdasarkan hasil percobaan penyusunan huruf pada Tabel 3, sistem menunjukkan performa yang cukup baik dalam mengenali dan menyusun huruf berdasarkan gesture bahasa isyarat SIBI. Pada kalimat tersebut terdapat 2 huruf yang salah dikenali, yaitu huruf V terdeteksi sebagai K, dan huruf S terdeteksi

sebagai I. Dari total 15 huruf dalam kalimat, sistem berhasil mengenali 13 huruf dengan benar, sehingga menghasilkan akurasi sebesar 86,6%. Kesalahan ini kemungkinan besar disebabkan oleh kemiripan visual antara huruf-huruf tertentu misalnya V dengan K, serta S dengan I, variasi sudut tangan, serta kondisi pencahayaan atau latar belakang saat pengambilan gesture. Hal ini berdampak langsung pada kalimat yang dihasilkan, karena kesalahan deteksi satu huruf saja dapat mengubah makna kata atau menurunkan keterbacaan kalimat.

Meskipun demikian, sistem tetap mampu menyusun kalimat utuh yang mendekati target, menunjukkan bahwa precision dan recall bekerja bersama dalam menjaga keseimbangan antara akurasi huruf yang dikenali dan kelengkapan huruf yang terdeteksi. Untuk meningkatkan performa sistem di masa depan, perlu dilakukan penambahan variasi data latih, optimasi parameter model, serta penggunaan teknik augmentasi untuk memperbaiki generalisasi terhadap kondisi nyata.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa sistem yang dibuat menggunakan framework *You Only Look Once (YOLO)* mampu mengenali pola huruf dari bahasa isyarat SIBI dengan performa yang cukup baik. Model yang dikembangkan menunjukkan rata-rata *precision* sebesar 0.835, *recall* sebesar 0.928, *accuracy* sebesar 0.831, dan *mAP@50* sebesar 0.968. Selain itu, sistem juga berhasil menyusun huruf-huruf hasil deteksi menjadi kalimat teks secara otomatis melalui input kamera secara *realtime*, dengan akurasi rata-rata mencapai 86,6%. Hal ini membuktikan bahwa sistem tidak hanya mampu mendeteksi huruf dari citra tangan pengguna, tetapi juga dapat merangkai huruf-huruf tersebut menjadi kalimat yang utuh dan bermakna.

Dengan demikian, tujuan dari penelitian ini telah tercapai, dan penerapan YOLOv11 terbukti efektif dalam mengenali dan menyusun huruf bahasa isyarat SIBI secara *realtime*. Meski begitu, sistem masih memiliki ruang untuk pengembangan lebih lanjut, khususnya dalam meningkatkan variasi data dan membedakan gesture huruf yang serupa agar sistem dapat diterapkan lebih optimal di dunia nyata.

Daftar Rujukan

- [1] Firsty Aufirandra, Bunga Adelya, and Syifa Ulfah, 2017. "Komunikasi Mempengaruhi Tingkah Laku individu," *JPGI (Jurnal Penelit. Guru Indones.*, vol. 2, no. 2, pp. 9–15.
- [2] V. Kissya, 2022. "Penggunaan Bahasa Isyarat Dalam Komunikasi Antara Penyandang Tunga Rungu, Guru, Serta Keluarga di (Sekolah Luar Biasa Pelitah Kasih) Rumah Tiga Ambon," *Hipotesa*, vol. 16, no. 1, pp. 20–34.
- [3] E. Zulpicha, 2018. "Konflik Kebijakan Penggunaan Sistem Bahasa Isyarat Indonesia Di Lingkungan Pendidikan Formal," *J. Anal. Sociol.*, vol. 6, no. 1, doi: 10.20961/jas.v6i1.18190.
- [4] R. Fatmawati, R. Asmara, Y. R. Prayogi, and R. Y. Hakkun,

2022. "Aplikasi Pembelajaran Sistem Isyarat Bahasa Indonesia (SIBI) Berbasis Voice Menggunakan OpenSIBI," *Technomedia J.*, vol. 7, no. 1, pp. 22–39, doi: 10.33050/tmj.v7i1.1690.
- [5] S. Allam, A. Balaram, Y. V. R. Rao, A. Kiran, H. Valpadasu, and S. Navya, 2024. "Sign Language Recognition Using Cnn," *Proc. 2nd Int. Conf. Adv. Smart, Secur. Intell. Comput. ASSIC 2024*, vol. 12, no. 12, doi: 10.1109/ASSIC60049.2024.10507969.
- [6] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González, 2023. "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Mach. Learn. Knowl. Extr.*, vol. 5, no. 4, pp. 1680–1716, doi: 10.3390/make5040083.
- [7] E. Altirika and W. P. Sari, 2023. "Pengembangan Deteksi Realtime untuk Bahasa Isyarat Indonesia dengan Menggunakan Metode Deep Learning Long Short Term Memory dan Convolutional Neural Network," *J. Teknol. Inform. dan Komput.*, vol. 9, no. 1, pp. 1–13, doi: 10.37012/jtik.v9i1.1272.
- [8] A. Kaisul Imam, K. Aeni, J. Raya Pagojengan Km, J. Tengah, and K. Kunci, 2023. "Deteksi Sistem Isyarat Bahasa Indonesia (Sibi) Menggunakan Algoritma Yolov5S,".
- [9] Agus Nugroho, R. Setiawan, A. Harris, and Beny, 2023. "Deteksi Bahasa Isyarat Bisindo Menggunakan Metode Machine Learning," *J. Process.*, vol. 18, no. 2, pp. 152–158, doi: 10.33998/processor.2023.18.2.1380.
- [10] B. K. Pratama, Sri Lestanti, and Yusniarsi Primasari, 2024. "Implementasi Algoritma You Only Look Once (YOLO) untuk Mendeteksi Bahasa Isyarat SIBI," *ProTekInfo(Pengembangan Ris. dan Obs. Tek. Inform.)*, vol. 11, no. 2, pp. 7–14, doi: 10.30656/protekinf.v11i2.9105.
- [11] A. Farizi *et al.*, 2024. "Implementasi Algoritma MobileNetV2 Berbasis CNN untuk Sistem Pengenalan Bahasa Isyarat SIBI," *BINA Insa. ICT J.*, [Online]. Available: <https://api.semanticscholar.org/CorpusID:278017429>
- [12] D. Permana and J. Sutopo, 2023. "APLIKASI PENGENALAN ABJAD SISTEM ISYARAT BAHASA INDONESIA (SIBI) DENGAN ALGORITMA YOLOv5 MOBILE APPLICATION ALPHABET RECOGNITION OF INDONESIAN LANGUAGE SIGN SYSTEM (SIBI) USING YOLOv5 ALGORITHM," *J. SimanteC*, vol. 11, no. 2, pp. 231–240.
- [13] B. Agung Saputra, F. Eka Putra, S. Julio Elsada Lattu, V. Handrianus Pranatawijaya, and R. Priskila, 2024. "IMPLEMENTASI ALGORITMA YOLO VERSI 8 UNTUK MEMBACA BAHASA ISYARAT Universitas Palangka Raya," *J. Multidisiplin Inov.*, vol. 8, no. 4, pp. 2246–6110.
- [14] K. Aeni and A. S. Millah, 2025. "Implementasi Deteksi Objek Dengan Model YOLOV8 pada Pengenalan Bahasa Isyarat Implementation of Object Detection with YOLOV8 Model in Sign Language Recognition," vol. 14, no. 105.
- [15] J. Zophie and H. H. Triharminto, 2020. "Implementasi Algoritma You Only Look Once (YOLO) menggunakan Web Camera untuk Mendeteksi Objek Statis dan Dinamis Implementation of You Only Look Once (YOLO) Algorithm using Web Camera for Static dan Dinamic Object Detection," vol. 1, no. 1, pp. 98–109.
- [16] Z. R. Mair and M. A. Rahmanda, 2025. "Perbandingan Versi Terbaik YOLO Dalam Mendeteksi Jarak Spasi Antar Baris Tulisan Tangan," *J. Sains, Nalar, dan Apl. Teknol. Inf.*, vol. 4, no. 2, pp. 103–110, doi: 10.20885/snati.v4.i2.40414.
- [17] Z. R. Mair and M. H. Irfani, 2023. "Permainan INGBAS (Gunting, Batu, Kertas) Menggunakan Arsitektur Convolutional Neural Network," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 1019–1026.
- [18] R. Heriansyah, D. A. Verano, Z. R. Mair, and others, 2024. "DETEKSI PENYAKIT DIABETES RETINOPATHY MENGGUNAKAN CITRA DIGITAL DENGAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN)," *Pros. SNAST*, pp. 311–320.
- [19] Zaid Romegar Mair *et al.*, 2023. "Z. R. Mair, W. Cholil, E. Yulianti, D. Marcelina, Theresiawati and I. N. Isnainiyah, 'Convolutional Neural Network Analysis on Handwriting Patterns and Its Relationship to Personality: A Systematical Review,' 2023 International Conference on Informatics, ," pp. 308–312.