

# Indonesian Sign Language Classification Using You Only Look Once

1<sup>st</sup> Dicky Luthfy  
Fakultas Teknik Elektro  
Universitas Telkom  
Bandung, Indonesia  
frdluthfy@student.telkomuniversity.ac.id

2<sup>nd</sup> Casi Setianingsih  
Fakultas Teknik Elektro  
Universitas Telkom  
Bandung, Indonesia  
setiacasie@telkomuniversity.ac.id

3<sup>rd</sup> Marisa W. Paryasto  
Fakultas Teknik Elektro  
Universitas Telkom  
Bandung, Indonesia  
marisaparyasto@telkomuniversity.ac.id

**Abstrak**— *Seiring majunya teknologi di bidang kamera digital, semakin banyak lapisan masyarakat yang terbantu oleh perkembangan teknologi tersebut, namun sayang ada beberapa kelompok masyarakat yang tidak dapat menikmati kemajuan tersebut seperti kaum disabilitas terkhusus Tuli dan Bisu. Tujuan sistem ini adalah untuk membantu kaum-kaum disabilitas tersebut agar dapat lebih mudah berkomunikasi dengan masyarakat umum melalui bahasa isyarat. Sistem yang dikembangkan dengan metode YOLOv5 dan menggunakan model pre-trained YOLOv5s untuk mengurangi waktu pelatihan. Model kemudian akan digunakan untuk melatih kelas-kelas baru dengan konfigurasi baru. Model yang sudah dilatih dengan konfigurasi tersebut kemudian akan digunakan untuk mengklasifikasikan 26 alfabet dari Sistem Bahasa Isyarat Indonesia atau biasa disingkat BISINDO. Pengujian sistem ini dilakukan berdasarkan beberapa skenario seperti jarak kamera, latar belakang pengambilan video dan tingkat pencahayaan area. Luaran yang didapatkan dari penelitian Tugas Akhir ini adalah sistem dapat mendeteksi 26 alfabet bahasa isyarat BISINDO secara real-time tanpa dipengaruhi oleh latar belakang dan tingkat pencahayaan tetapi dipengaruhi oleh jarak kamera dan objek. Hasil konfigurasi performa terbaik pada penelitian ini adalah dataset dengan distribusi 70% data training; 20% data validation; 10% data testing, 300 epochs, 16 batch size, dan 0.01 learning rate yang menghasilkan nilai mAP@0.5IoU sebesar 99.27%.*

**Kata kunci**— *BISINDO, disabilitas, YOLO*

## I. PENDAHULUAN

Bahasa merupakan alat utama berkomunikasi dan sudah menjadi faktor penting dalam kehidupan sehari-hari. Selain bahasa yang sudah dikenal secara umum dan dipraktikkan oleh banyak bagian masyarakat, terdapat juga bahasa yang acap digunakan oleh masyarakat-masyarakat yang memiliki disabilitas seperti contohnya bahasa isyarat, namun belum banyak terdapat metode yang memperkenankan masyarakat yang tidak mengerti bahasa isyarat untuk berkomunikasi dengan masyarakat Bisu atau Tuli. Merujuk pada data Sistem Informasi Manajemen Penyandang Disabilitas dari Kementerian Sosial pada tahun 2019 tentang persentase penyandang disabilitas di Indonesia berdasarkan jenis, Tuli mencakup 7,03% diantaranya atau sekitar 637.535 orang [1].

Berdasar pada keadaan tersebut, diajukanlah sebuah solusi sistem yang dapat membaca bahasa

isyarat kemudian mentranslasikannya ke alfabet bahasa Indonesia yang dapat dimengerti oleh mayoritas masyarakat. Solusi ini dipilih karena perlunya ada metode berkomunikasi yang lebih inklusif terhadap mayoritas kelompok masyarakat. Sebelumnya sudah ada penelitian serupa tetapi terbatas pada kemampuan sistem yang hanya bisa dijalankan melalui perangkat komputer, pada penelitian ini akan digunakan YOLOv5 untuk algoritma pembantu dan penerapannya pada aplikasi mobile untuk memudahkan penggunaan sistem. Solusi ini dapat membantu masyarakat untuk belajar bahasa isyarat.

## II. KAJIAN TEORI

### A. Bahasa Isyarat

Bahasa Isyarat adalah salah satu metode berkomunikasi dengan orang Tuli. Bahasa Isyarat diungkapkan melalui gerakan-gerakan manual seperti gerakan tangan dan kepala, maupun non-manual seperti pandangan mata dan gestur-gestur arah gerakan tubuh [2]. Selayaknya beragam bahasa yang ada di dunia, bahasa isyarat juga memiliki variasi-variasinya sendiri yang bergantung dengan struktur tatabahasa masing-masing negara walaupun kedua negara memiliki bahasa yang sama seperti sistem isyarat American Sign Language dan British Sign Language yang sama-sama berdasarkan bahasa Inggris tetapi memiliki tatabahasa yang berbeda [3]. Meskipun begitu, tidak berarti semua bahasa isyarat tidak memiliki kesamaan sama sekali, seperti contohnya sistem isyarat Indonesia yang memiliki kesamaan dengan sistem bahasa isyarat Bangladesh.

### B. Bahasa Isyarat Indonesia

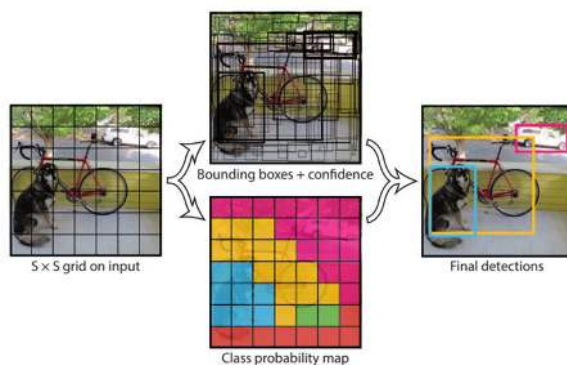
Bahasa Isyarat Indonesia atau BISINDO adalah sistem bahasa isyarat yang umum digunakan oleh Teman Tuli maupun pengguna bahasa isyarat meskipun bahasa yang diakui secara resmi oleh pemerintah adalah Sistem Isyarat Bahasa Indonesia (SIBI). BISINDO lebih umum digunakan karena sistem bahasa isyarat ini dikembangkan oleh Teman Tuli itu sendiri [4]. Sistem BISINDO memiliki tatanan sistematis yang diekspresikan dengan isyarat jari, tangan, gestur tubuh, dan gerakan-gerakan yang melambangkan kosakata dalam Bahasa Indonesia. Alfabet-alfabet BISINDO terdiri dari bentuk-bentuk

gerakan tangan statis dengan pengecualian alfabet J dan R yang berbentuk gestur.

### C. You Only Look Once

You Only Look Once atau biasa disingkat YOLO adalah metode pendeteksian objek yang populer dikarenakan kecepatannya dalam memproses citra. Berbeda dengan CNN dan turunan CNN lainnya, YOLO hanya menggunakan satu lapisan jaringan untuk memprediksi objek apa yang ada dalam gambar atau video yang mengakibatkan lebih tingginya kecepatan YOLO jika dibandingkan dengan CNN dengan konsekuensi akurasi yang sedikit lebih rendah [5]. Pada penelitian ini, versi YOLO yang digunakan adalah YOLOv5 atau biasa disebut juga ultralytics-yolo merupakan versi “informal” dari YOLOv4. Meskipun menggunakan arsitektur yang sama, YOLOv5 berhasil memangkas ukuran model dari yang awalnya sebesar 245 MB menggunakan YOLOv4, menjadi hanya 27 MB dengan YOLOv5 dengan ukuran gambar yang sama. Meskipun begitu, YOLOv5 memiliki AP 36.7%, lebih rendah daripada YOLOv4 yang memiliki AP 41.2% pada konfigurasi dataset dan environment yang sama [6].

Secara garis besar, cara kerja YOLO dibagi lagi menjadi beberapa tingkat. Tingkat pertama yaitu citra yang akan diproses dibagi menjadi beberapa sel jaringan, jumlah sel jaringan dapat dihitung menggunakan cara membagi resolusi gambar dengan jumlah stride yang ditentukan agar menghasilkan kotak 8x8, 16x16, dan 32x32. Dikarenakan pada YOLO umumnya digunakan resolusi gambar 640x640, maka stride yang digunakan adalah 80x80 untuk objek ukuran kecil, 40x40 untuk objek ukuran sedang, dan 20x20 untuk objek ukuran besar [7].



GAMBAR 1  
ALUR KERJA YOLO.

Masing-masing dari kotak pembatas memiliki 4 koordinat dan satu parameter pengukur yaitu  $x, y, w, h$ , dan ketepatan. Koordinat  $(x, y)$  digunakan untuk menandakan lokasi tengah kotak yang relatif dari letak kotak tersebut. Koordinat  $(w, h)$  digunakan untuk menandakan tinggi dan lebar dari gambar secara keseluruhan, sedangkan parameter ketepatan mengukur IoU (Intersection Over Union) atau biasa disebut rasio tumpang tindih yang menandakan area tumpang tindih antara kotak objek

yang akan dideteksi dan kotak objek kebenaran dasar [8].

Selain 4 koordinat dan satu parameter pengukur tadi, masing masing sel juga memiliki satu parameter lagi yaitu kelas dari objek yang dinotasikan dengan simbol  $c$ , yang hanya ada satu pada setiap sel. Pada YOLO, digunakan metode Non Maximum Suppression untuk mengeliminasi sel-sel yang mendeteksi objek yang sama dengan cara melihat objek sel yang memiliki keakuratan paling tinggi dan mengeliminasi sel-sel pembatas lainnya yang mendeteksi objek yang sama [9]. Masing-masing sel pada YOLO hanya dapat memprediksi satu jenis kelas, yang berujung kepada kelemahan metode ini yaitu tidak dapat mengklasifikasi dua objek berbeda jika mereka berada dalam satu sel yang sama [10].

### D. Transfer Learning

Transfer learning adalah metode yang menggunakan pretrained weight dari model algoritma yang telah dilatih sebelumnya untuk mempersingkat waktu pelatihan dengan tidak mengorbankan akurasi. Informasi tersebut kemudian digunakan lagi untuk melatih model baru dengan parameter, konfigurasi, dan dataset yang berbeda [11]. Model YOLOv5 yang digunakan pada penelitian ini adalah YOLOv5s yang sebelumnya dilatih untuk mengenali 80 kelas pada MS COCO dataset. Untuk mencapai tujuan penelitian, dilakukan fine tuning pada model YOLOv5s agar hanya dapat mendeteksi 26 kelas alfabet BISINDO.

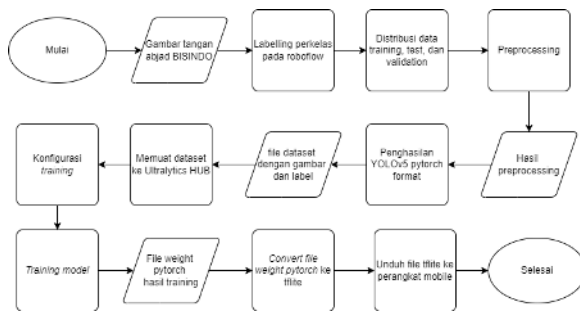
### E. Parameter Performa

Parameter performa adalah parameter-parameter yang didapatkan setelah melakukan proses training. Nilai-nilai dari parameter ini kemudian akan digunakan untuk menilai keoptimalan hasil training. Parameter-parameter yang digunakan dalam penelitian ini adalah *Precision*, *Recall*, *F1 Score*, dan *mAP@0.5IoU*

## III. METODE

### A. Arsitektur Aplikasi

Pada arsitektur aplikasi deteksi bahasa isyarat, langkah pertama yang dilakukan adalah mengambil dataset gambar bahasa isyarat BISINDO dan kemudian memberi label per-kelas pada masing masing gambar tersebut. Setelah proses label selesai, seluruh gambar akan dibagi menjadi 3 set yaitu *training*, *test*, dan *validation*. Setelah melewati tahap *preprocessing*, data kemudian akan diekstrak kedalam format YOLOv5 Pytorch. Langkah selanjutnya adalah memuat data-data yang sudah dilabel tadi ke dalam Ultralytics HUB untuk dilakukan proses training menggunakan Google Collab. Jika training sudah selesai, Ultralytics HUB akan mengubah file weight pytorch menjadi tflite dan memuatnya ke aplikasi mobile.



GAMBAR 2  
ARSITEKTUR APLIKASI.

## B. Analisa Kebutuhan Sistem

Berdasarkan dengan penjelasan gambaran umum sistem, dapat diketahui sistem deteksi bahasa isyarat membutuhkan beberapa aspek yang dapat mendukung implementasi sistem dalam bentuk *mobile*. Aspek yang dibutuhkan, yaitu Data, Perangkat Lunak, dan Kebutuhan Pengguna.

### 1. Kebutuhan Data

Data yang digunakan untuk kebutuhan sistem adalah data 30 gambar masing masing alfabet sejumlah 26 huruf yang membuat total data gambar berjumlah 780 gambar. Pemotretan foto tangan menggunakan kamera yang menghasilkan gambar berukuran 1:1 dengan background bervariasi.



GAMBAR 3  
CONTOH DATASET.

Untuk mencapai hasil *training* yang optimal, berbagai augmentasi dilakukan seperti pergeseran sudut pengambilan gambar dan sudut tangan. Foto-foto tersebut kemudian diproses sebagaimana berikut:

#### a. Anotasi

Setelah pengambilan dataset, setiap data perlu diberi label sesuai dengan format algoritma yang digunakan. Pada pelabelan YOLO, setiap gambar diberi file text (.txt) tambahan yang memiliki nama sama dengan gambar. Masing masing *file text* (.txt) berisi anotasi *object class*, *object coordinate*, *height*, dan *width*.

#### b. Preprocessing dan augmentasi data.

Sebelum data diolah oleh model lebih lanjut, data perlu diolah terlebih dahulu agar tercapai hasil yang optimal dan tidak terjadi overfitting. Preprocessing diaplikasikan ke seluruh gambar sedangkan augmentasi hanya diaplikasikan ke gambar training.

TABEL 1  
TABEL PREPROCESSING DAN AUGMENTASI DATA

Preprocessing	Augmentasi
<i>Auto-orient</i>	<i>Flip Horizontal</i>
<i>Resize to 640x640</i>	<i>Crop 0% Minimum Zoom, 20% Maximum Zoom</i>
	<i>Rotation Between -5° and +5°</i>
	<i>Shear ±5° Horizontal, ±5° Vertical</i>
	<i>Grayscale on 10% of images</i>
	<i>Brightness Between -25% and +25%</i>
	<i>Blur up to 1.25px</i>

### c. Pembagian data *training*, *validation*, dan *test*.

Data *training* adalah data yang digunakan algoritma untuk mengklasifikasikan dan melatih model untuk mendeteksi kelas kelas baru yang belum pernah dilihat sebelumnya. Data ini biasanya memiliki rasio paling tinggi di antara ketiga set data. Data *validation* adalah data yang digunakan saat *training* yang berfungsi untuk mengoreksi nilai error. Sedangkan data *test* digunakan ketika model sudah selesai dilatih dan digunakan untuk membandingkan sebgas apa performa model dari hasil *training* yang dilakukan.

### 2. Perangkat Lunak

Perangkat lunak yang digunakan dalam perancangan model dan pembuatan aplikasi *mobile* menggunakan perangkat lunak dengan spesifikasi sebagai berikut:

- Bahasa pemrograman Python versi 3.10
- Sistem operasi windows 10 Pro
- Pengelola dataset Roboflow
- Pengelola GPU *computing* CUDA versi 11.3.
- Android Studio

### 3. Kebutuhan Pengguna

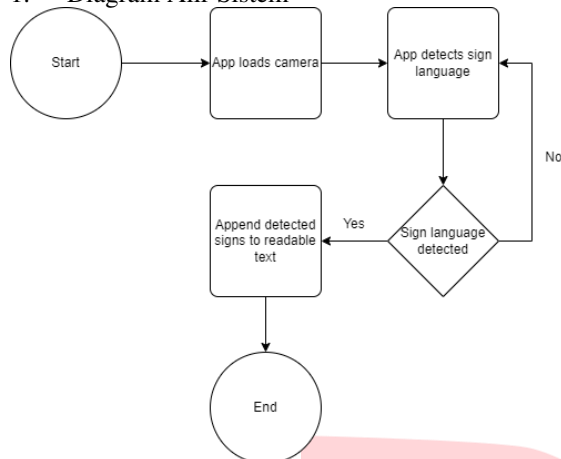
Pengguna yang akan menggunakan sistem penjadwalan perangkat listrik harus memenuhi kriteria sebagai berikut:

- Teman Tuli dan Bisu.
- Individu lain yang ingin berkomunikasi dengan bahasa isyarat.
- Individu lain yang bisa berkomunikasi dengan bahasa isyarat.

### C. Perancangan Sistem

Perancangan sistem adalah gambaran dari sistem deteksi bahasa isyarat yang akan dirancang Dalam perancangan sistem ini terdapat gambaran diagram alir sistem dan proses pelatihan model YOLOv5.

### 1. Diagram Alir Sistem

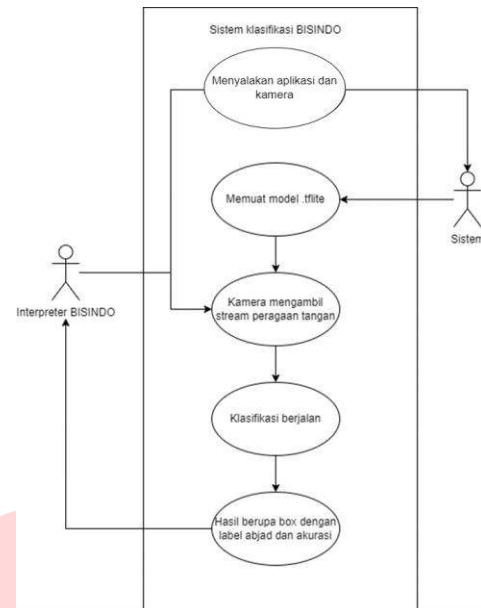


GAMBAR 4  
DIAGRAM ALIR SISTEM.

Diagram alir sistem pada gambar menunjukkan alur kerja dari sistem deteksi bahasa isyarat. Langkah pertama setelah aplikasi dijalankan adalah aplikasi akan membuka kamera yang berjalan paralel dengan pemuatan *file weight* hasil *training* dari algoritma. Setelah aplikasi berhasil dibuka, pengguna melakukan gerakan bahasa isyarat di depan kamera. Jika gerakan yang ditunjukkan sesuai dengan bahasa isyarat BISINDO, layar akan menunjukkan kelas hasil deteksi beserta kotak pembatas dan *confidence score*-nya. Setelah berhasil mendeteksi kelas alfabet, pengguna bisa memilih untuk memasukkan huruf yang terdeteksi kedalam rangkaian kalimat dengan menekan tombol “ADD” pada aplikasi, dan membersihkan rangkaian kalimat yang sudah terbuat dengan menekan tombol “CLEAR”.

### 2. Use Case Diagram

Berdasarkan use case diagram pada gambar 5, diperagakan penggunaan sistem klasifikasi bahasa isyarat. Pada sistem ini, terdapat dua aktor yaitu yang pertama Interpreter BISINDO sebagai aktor yang memperagakan gerakan abjad BISINDO dan yang kedua Sistem sebagai aktor yang memegang kendali penuh dalam sistem mulai dari inferensi gambar hingga menampilkan gambar objek yang sudah diklasifikasi.



GAMBAR 5  
USE CASE DIAGRAM.

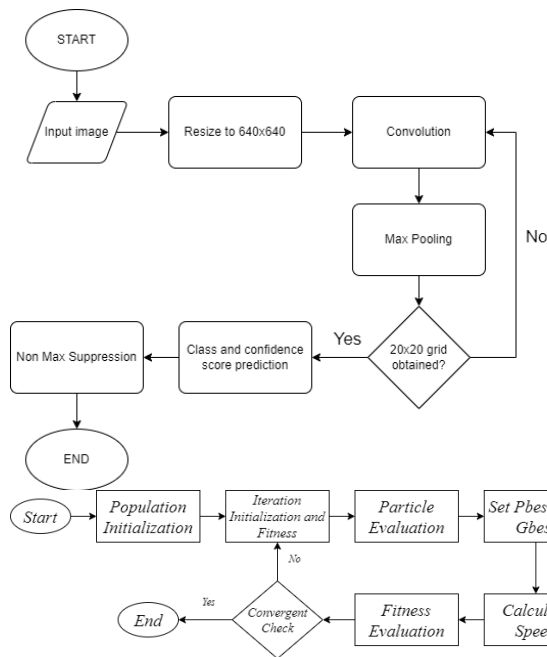
### 3. Proses Algoritma YOLO

Pada algoritma yolo, sebelum diproses, ukuran gambar akan diubah terlebih dahulu menjadi 640x640 dan diberi bounding box kebenaran dasar sejumlah SxS. Setelah membagi gambar menjadi SxS bagian, dilakukan penentuan bounding box object yang ditandai dengan vektor persamaan 3.1.

$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c \end{bmatrix} \quad (3.1)$$

Dengan  $p_c$  adalah tingkat kemungkinan *bounding box* tersebut mengandung objek,  $b_x$  dan  $b_y$  adalah koordinat titik tengah *bounding box* relatif terhadap keseluruhan gambar,  $b_h$   $b_w$  sebagai panjang dan lebar dari *bounding box*, dan  $c$  sebagai kemungkinan *bounding box* tersebut mengandung kelas tertentu. Pada penelitian ini, terdapat 26 kelas klasifikasi, maka dari itu persamaan 3.1 akan memiliki nilai  $c$  sebanyak 26, mulai dari  $c_1$  hingga  $c_{26}$ . Setelah mendapatkan nilai vektor  $y$  dari semua *bounding box*, akan dilakukan proses konvolusi untuk memperkecil ukuran gambar dengan cara mengalikan informasi masing masing nilai vektor  $y$  dengan nilai kernel atau nilai filter proses konvolusi, yang kemudian akan menghasilkan gambar dengan nilai citra baru.



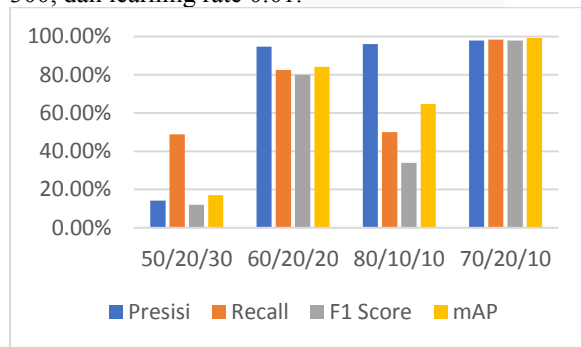


GAMBAR 6  
FLOWCHART ALGORITMA YOLO

#### IV. HASIL DAN PEMBAHASAN

##### A. Pengujian Distribusi Dataset

Berikut pengujian distribusi dataset dilakukan untuk mengetahui nilai rasio dataset terbaik untuk konfigurasi sistem. Konfigurasi awal yang digunakan untuk pengujian ini adalah dengan batch size 16, epoch 300, dan learning rate 0.01.

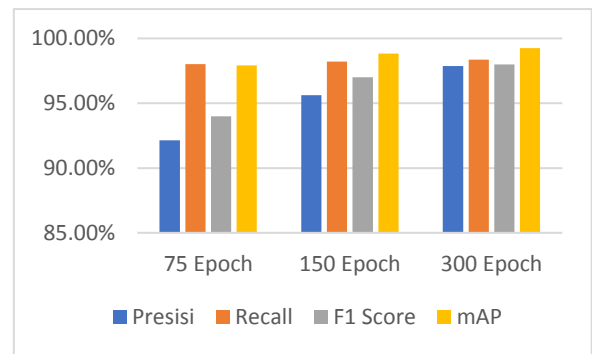


GAMBAR 7  
GRAFIK PENGUJIAN DISTRIBUSI DATASET.

Pengujian distribusi dataset pada grafik dapat disimpulkan bahwa pembagian dataset dengan rasio 70/20/10 menghasilkan hasil yang paling bagus.

##### B. Pengujian Epoch

Berikut pengujian epoch dilakukan untuk mengetahui nilai epoch terbaik untuk konfigurasi sistem. Konfigurasi yang digunakan untuk pengujian ini adalah dengan batch size 16, pembagian dataset 70/20/10, dan learning rate 0.01.



GAMBAR 8  
GRAFIK PENGUJIAN NILAI R1

Pengujian epoch pada grafik dapat disimpulkan bahwa 300 menghasilkan hasil yang paling bagus

##### C. Pengujian inferensi latar belakang

Pengujian ini dilakukan untuk mengetahui keberhasilan model yang sudah dilatih untuk menjalankan inferensi dengan berbagai latar.

TABEL 2  
TABEL PENGUJIAN LATAR BELAKANG.

Latar belakang	Nilai aktual	Nilai terdeteksi	Akurasi
Putih	77	78	98.7%
Bermotif	77	78	98.7%

Dapat disimpulkan bahwa model berjalan dengan baik dan tidak dipengaruhi oleh motif latar belakang.

##### D. Pengujian inferensi jarak objek

Pengujian ini dilakukan untuk mengetahui keberhasilan model yang sudah dilatih untuk mendeteksi objek dengan berbagai jarak.

TABEL 3  
TABEL PENGUJIAN JARAK OBJEK

Jarak	Nilai aktual	Nilai terdeteksi	Akurasi
50cm	2	78	2%
25cm	76	78	98.7%

Dapat disimpulkan bahwa model tidak dapat mendeteksi objek yang melebihi 25cm dari kamera.

##### E. Pengujian inferensi tingkat pencahayaan

Pengujian ini dilakukan untuk mengetahui keberhasilan model yang sudah dilatih untuk menjalankan inferensi di berbagai tingkat pencahayaan.

TABEL 4  
TABEL PENGUJIAN TINGKAT PENCAHAYAAN.

Tingkat pencahayaan	Nilai aktual	Nilai terdeteksi	Akurasi
10 Lux	77	78	98.7%
700 Lux	77	78	98.7%
20.000 Lux	77	78	98.7%

Dapat disimpulkan bahwa model berjalan dengan baik dan tidak dipengaruhi oleh tingkat pencahayaan.

## V. KESIMPULAN

Berdasarkan hasil dari pengujian dan analisis yang telah dilakukan pada Tugas Akhir ini, algoritma YOLOv5 dengan 780 gambar awal dataset sebelum diberi augmentasi (30 gambar per-kelas) memiliki hasil pelatihan paling optimal pada pembagian dataset 70%;20%;10% dan epoch 300. Hasil pengujian sistem menandakan bahwa model tidak terpengaruh oleh latar belakang dan tingkat pencahayaan hingga 20000 lux, tetapi model tidak dapat mendeteksi objek jika jarak antar kamera dan objek >25cm.

## REFERENSI

- [1] K. Snoddon, "Action Research with a Family ASL Literacy Program," *Writing & Pedagogy*, vol. 3, no. 2, Dec. 2011, doi: 10.1558/WAP.V3I2.265.
- [2] H. Cooper, B. Holt, and R. Bowden, "Sign Language Recognition," *Visual Analysis of Humans*, pp. 539–562, 2011, doi: 10.1007/978-0-85729-997-0\_27.
- [3] "Sistem Informasi Penyandang Disabilitas - Kementerian Sosial RI." <http://simpd.kemensos.go.id/> (accessed Aug. 10, 2022).
- [4] X. Huang *et al.*, "PP-YOLOv2: A Practical Object Detector," Apr. 2021, doi: 10.48550/arxiv.2104.10419.
- [5] R. Fatmi, S. Rashad, R. Integlia, and G. Hutchison, "American Sign Language Recognition using Hidden Markov Models and Wearable Motion Sensors," *Trans. Mach. Learn. Data Min.*, vol. 10, pp. 41–55, 2017.
- [6] P. Patel and N. Patel, "Vision Based Real-time Recognition of Hand Gestures for Indian Sign Language using Histogram of Oriented Gradients Features," *International Journal of Next-Generation Computing*, pp. 92–102, Jul. 2019, doi: 10.47164/IJNGC.V10I2.158.
- [7] A. Corovic, V. Ilic, S. Duric, M. Marijan, and B. Pavkovic, "The Real-Time Detection of Traffic Participants Using YOLO Algorithm," *2018 26th Telecommunications Forum, TELFOR 2018 - Proceedings*, 2018, doi: 10.1109/TELFOR.2018.8611986.
- [8] "What is Sign Language?", Accessed: Aug. 10, 2022. [Online]. Available: <http://www.lsadc.org>
- [9] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A Forest Fire Detection System Based on Ensemble Learning," *Forests 2021, Vol. 12, Page 217*, vol. 12, no. 2, p. 217, Feb. 2021, doi: 10.3390/F12020217.
- [10] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/S40537-019-0197-0/FIGURES/33.