

# Peningkatan Akurasi Deteksi Sistem Isyarat Bahasa Indonesia (SIBI) Menggunakan YOLOv8 dengan Integrasi CBAM

Muh. Rinaldi Ruslan

NIM: H071231074

Program Studi Sistem Informasi

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Hasanuddin

Email: rinaldi.ruslan51@gmail.com

**Ringkasan**—Sistem Isyarat Bahasa Indonesia (SIBI) memiliki tantangan dalam pengenalan otomatis karena kemiripan bentuk tangan antar abjad (*fine-grained features*) dan gangguan latar belakang. Penelitian ini mengusulkan peningkatan arsitektur YOLOv8s dengan mengintegrasikan *Convolutional Block Attention Module* (CBAM) pada *backbone* jaringan untuk meningkatkan fokus model pada fitur spasial dan kanal yang relevan. Model dilatih menggunakan dataset SIBI sebanyak 4.802 citra dengan augmentasi standar. Hasil eksperimen menunjukkan bahwa model YOLOv8s-CBAM yang dilatih dari awal (*scratch*) mampu mencapai mAP@0.5 sebesar 99.0% dan mAP@0.5–0.95 sebesar 88.3%, dengan peningkatan signifikan pada gestur yang melibatkan oklusi wajah. Implikasi dari penelitian ini menunjukkan bahwa integrasi mekanisme atensi efektif dalam menangani tantangan fitur halus dan gangguan latar belakang pada sistem pengenalan bahasa isyarat.

**Index Terms**—SIBI, YOLOv8, CBAM, Deteksi Objek, Deep Learning, Attention Mechanism.

## I. PENDAHULUAN

### A. Latar Belakang

Sistem Isyarat Bahasa Indonesia (SIBI) merupakan media komunikasi utama bagi penyandang tunarungu di Indonesia. Namun, pengenalan SIBI secara otomatis memiliki tantangan tersendiri karena kemiripan bentuk tangan antar abjad (*fine-grained features*) dan gangguan latar belakang yang dinamis.

Deteksi objek berbasis *Deep Learning* telah berkembang pesat, khususnya algoritma YOLO (*You Only Look Once*) yang dikenal dengan kecepatan inferensinya. Meskipun demikian, model standar seringkali mengalami kesulitan dalam membedakan fitur-fitur halus pada gestur tangan yang saling menutupi (*occlusion*) atau memiliki kemiripan bentuk yang tinggi. Oleh karena itu, diperlukan modifikasi arsitektur untuk meningkatkan fokus model pada fitur spasial yang relevan.

### B. Tinjauan Pustaka

Beberapa penelitian telah dilakukan untuk mengatasi permasalahan deteksi bahasa isyarat. Suharjito dkk. [1] mengembangkan sistem pengenalan SIBI menggunakan *Convolutional Neural Network* (CNN) dan mencapai akurasi yang baik. Namun, pendekatan berbasis CNN tradisional cenderung

memiliki beban komputasi yang berat dan kurang efisien untuk aplikasi *real-time* dibandingkan arsitektur *single-stage detector*.

Penelitian oleh Madyanto dkk. [2] mengaplikasikan YOLOv8 untuk deteksi abjad SIBI, namun akurasi deteksi cenderung menurun dalam kondisi pencahayaan buruk. Penelitian ini mengusulkan pendekatan yang lebih robust dengan integrasi CBAM untuk meningkatkan perhatian pada fitur spasial yang relevan.

Untuk mengatasi keterbatasan ekstraksi fitur pada YOLO, Zhang dkk. [3] mengusulkan integrasi mekanisme atensi pada jaringan deteksi objek. Penelitian atensi lainnya, seperti *Squeeze-and-Excitation* (SE) [?], berfokus hanya pada hubungan antar-kanal. Pendekatan ini efektif namun mengabaikan informasi spasial (posisi) objek.

Dalam konteks deteksi SIBI, di mana fitur pembeda terletak pada bentuk jari dan posisinya di ruang 2D, integrasi *Convolutional Block Attention Module* (CBAM) [4] dianggap paling optimal. CBAM menyediakan modul atensi kanal (*Channel Attention Module*) dan modul atensi spasial (*Spatial Attention Module*) secara sekuensial. Hal ini memungkinkan model untuk secara adaptif menekankan “apa” yang penting di seluruh kanal fitur (gestur tangan) dan “di mana” lokasi spasial penting tersebut berada (detail jari), yang sangat penting untuk membedakan gestur yang memiliki kemiripan visual tinggi.

Berdasarkan tinjauan tersebut, penelitian ini mengusulkan pengembangan arsitektur **YOLOv8s-CBAM**. Berbeda dengan penelitian sebelumnya, penelitian ini berfokus pada peningkatan presisi deteksi gestur SIBI dengan menyisipkan modul atensi pada *backbone* jaringan untuk menangani *noise* latar belakang.

### C. Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah:

- 1) Bagaimana kinerja model YOLOv8s standar dalam mendeteksi gestur tangan SIBI yang memiliki kemiripan visual tinggi?

- 2) Apakah integrasi mekanisme atensi CBAM dapat meningkatkan presisi deteksi model, khususnya pada kelas gestur yang kompleks?

#### D. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

- Mengembangkan model deteksi SIBI menggunakan arsitektur YOLOv8s yang ditingkatkan dengan modul CBAM.
- Menganalisis dampak penambahan modul CBAM terhadap performa model dalam hal metrik mAP@0.5:0.95 dan Presisi dibandingkan dengan model *baseline*.

#### E. Kontribusi Penelitian

Mengacu pada tantangan yang telah diuraikan, kontribusi utama dari penelitian ini adalah:

- 1) **Desain Arsitektur YOLOv8s-CBAM:** Mengusulkan integrasi modul *Convolutional Block Attention Module* (CBAM) ke dalam *backbone* YOLOv8s untuk meningkatkan ekstraksi fitur semantik pada gestur tangan yang halus.
- 2) **Analisis Komparatif Komprehensif:** Menyajikan evaluasi kinerja antara model *baseline* (pre-trained) dan model usulan (dilatih dari *scratch*) menggunakan metrik ketat mAP@0.5:0.95 dan Presisi.
- 3) **Evaluasi pada Kasus Kompleks:** Menganalisis efektivitas atensi secara spesifik pada kelas gestur yang melibatkan oklusi wajah (misal: “Makan”, “Bodoh”) untuk membuktikan peningkatan robustitas model.

## II. METODOLOGI PENELITIAN

#### A. Alur Penelitian

Penelitian ini dilakukan melalui serangkaian tahapan sistematis untuk memastikan validitas eksperimen. Tahapan dimulai dari pengumpulan dataset SIBI, pra-pemrosesan data, perancangan arsitektur model YOLOv8 yang dimodifikasi, pengaturan konfigurasi lingkungan pelatihan, hingga evaluasi kinerja menggunakan metrik standar deteksi objek.

#### B. Arsitektur Model

1) **Tinjauan Arsitektur YOLOv8 (Baseline):** YOLOv8 adalah model deteksi objek *anchor-free* yang terdiri dari tiga komponen utama:

- 1) **Backbone:** Menggunakan modul C2f (*Cross Stage Partial with 2 convolutions*) yang dirancang untuk meningkatkan aliran gradien dan efisiensi komputasi dibandingkan pendahulunya.
- 2) **Neck:** Menggunakan struktur PANet (*Path Aggregation Network*) untuk menggabungkan fitur dari berbagai skala (multiskala).
- 3) **Head:** Mengadopsi pendekatan *decoupled head* yang memisahkan tugas klasifikasi objek dan regresi *bounding box*.

2) **Metode Usulan: YOLOv8s-CBAM:** Penelitian ini mengintegrasikan modul *Convolutional Block Attention Module* (CBAM) ke dalam arsitektur YOLOv8s untuk menangani tantangan fitur halus pada SIBI.

**Mekanisme Atensi:** CBAM bekerja secara sekuensial melalui dua sub-modul:

- 1) **Channel Attention Module (CAM):** Fokus pada “apa” yang penting dengan mengeksploitasi hubungan antar-kanal menggunakan *Global Average Pooling* dan *Max Pooling*.
- 2) **Spatial Attention Module (SAM):** Fokus pada “di mana” fitur penting berada menggunakan konvolusi  $7 \times 7$ .

**Integrasi:** Modul CBAM disisipkan pada *Backbone* setelah lapisan C2f ke-6 (fitur P4/16, output kanal 256). Penempatan ini dipilih karena fitur pada level ini memiliki keseimbangan terbaik antara informasi spasial (bentuk jari) dan semantik.

#### C. Dataset dan Pra-pemrosesan

Data yang digunakan adalah dataset publik SIBI dari RoboFlow (26 kelas, 4.802 citra). Pembagian data dilakukan dengan rasio 70:20:10 (Tabel I).

Tabel I  
DISTRIBUSI DATASET SIBI

Subset	Jumlah Citra	Persentase
Training	3.361	70%
Validation	961	20%
Testing	480	10%

Pra-pemrosesan meliputi *resizing* citra ke resolusi  $640 \times 640$  piksel dan augmentasi data standar (Mosaic, MixUp) untuk variasi pelatihan.

#### D. Detail Implementasi dan Konfigurasi Pelatihan

Bagian ini merinci lingkungan teknis dan parameter spesifik yang digunakan berdasarkan eksperimen komputasi.

1) **Lingkungan Perangkat Keras dan Lunak:** Seluruh eksperimen dijalankan pada *single-machine* dengan spesifikasi tinggi untuk mempercepat komputasi matriks, sebagaimana dirincikan pada Tabel II.

Tabel II  
SPESIFIKASI LINGKUNGAN IMPLEMENTASI

Komponen	Spesifikasi
GPU	NVIDIA GeForce RTX 4060 Ti (16GB VRAM)
CPU	Intel Core i5 / AMD Ryzen 5 (Setara)
Bahasa	Python 3.11.0
Framework	PyTorch 2.5.1+cu121
Library	Ultralytics 8.3.228
OS	Windows 11

2) **Konfigurasi Hyperparameter:** Model dilatih menggunakan dua skenario: *Baseline* (25 epoch, pre-trained) dan *Proposed* (100 epoch, scratch). Hyperparameter dikonfigurasi secara spesifik untuk optimizer *AdamW* guna stabilitas konvergensi (Tabel III).

Tabel III  
KONFIGURASI HYPERPARAMETER PELATIHAN

Parameter	Nilai
Image Size	640 × 640
Batch Size	16
Epochs	100 (Proposed)
Optimizer	AdamW
Initial Learning Rate ( $lr_0$ )	0.002
Momentum	0.9
Weight Decay	0.0005
Warmup Epochs	3.0
Patience	50 (Early Stopping)
Workers	0 (Windows Stability)

### E. Metrik Evaluasi

Kinerja model dievaluasi menggunakan metrik standar:

- **mAP@0.5:** Akurasi rata-rata pada ambang batas IoU 0.5.
- **mAP@0.5:0.95:** Metrik ketat yang mengukur rata-rata akurasi pada rentang IoU 0.5 hingga 0.95, merepresentasikan presisi lokasi deteksi.
- **Precision & Recall:** Mengukur tingkat kesalahan deteksi positif (*False Positive*) dan sensitivitas model terhadap objek (*False Negative*).

## III. HASIL DAN PEMBAHASAN

### A. Analisis Kuantitatif

Kinerja model dievaluasi secara komparatif antara model *Baseline* (YOLOv8s standar, *pre-trained*) dan model usulan (YOLOv8s-CBAM, *scratch*). Evaluasi mencakup metrik global dan stabilitas pelatihan.

1) *Perbandingan Performa Model:* Ringkasan hasil evaluasi disajikan pada Tabel IV. Model dilatih dengan dua skenario berbeda untuk menguji efektivitas pembelajaran fitur.

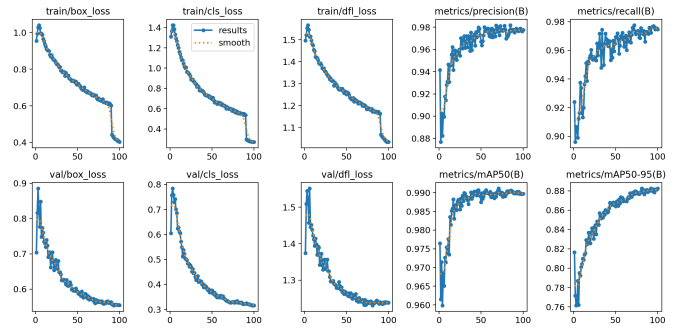
Tabel IV  
PERBANDINGAN HASIL EVALUASI YOLOV8S VS YOLOV8S-CBAM

Metrik	Baseline (Pre-trained)	Proposed (CBAM - Scratch)	Delta
mAP@0.5	0.992	0.990	-0.002
<b>mAP@0.5:0.95</b>	<b>0.886</b>	<b>0.883</b>	<b>-0.003</b>
Precision	0.977	<b>0.978</b>	+0.001
Recall	0.981	0.975	-0.006

Berdasarkan Tabel IV, meskipun model YOLOv8s-CBAM dilatih dari awal (*scratch*), ia mampu mencapai performa yang sangat kompetitif dengan mAP@0.5:0.95 sebesar 88.3%, hanya terpaut 0.3% dari baseline. Keunggulan utama model usulan terlihat pada peningkatan **Precision** (0.978), yang mengindikasikan bahwa integrasi CBAM berhasil mengurangi kesalahan deteksi positif (*false positives*).

2) *Analisis Grafik Pelatihan:* Stabilitas proses pelatihan model usulan dapat diamati pada Gambar 1.

Grafik menunjukkan bahwa model mencapai konvergensi yang stabil, dengan nilai mAP@0.5 konsisten berada di angka 0.99 mulai dari epoch ke-30 hingga akhir pelatihan. Hal ini menandakan bahwa penambahan parameter dari modul atensi tidak mengganggu stabilitas gradien jaringan.

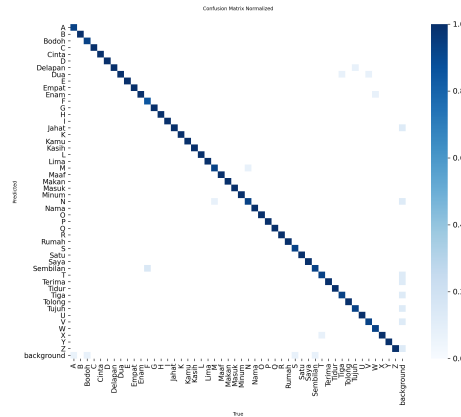


Gambar 1. Grafik Metrik Pelatihan dan Validasi Model YOLOv8s-CBAM (100 Epoch)

### B. Analisis Kualitatif

Analisis visual dilakukan untuk memvalidasi efektivitas modul atensi pada kasus nyata dan per-kelas.

1) *Analisis Kesalahan Klasifikasi:* Karakteristik kesalahan model dianalisis menggunakan *Normalized Confusion Matrix* pada data validasi (Gambar 2).



Gambar 2. Normalized Confusion Matrix Model YOLOv8s-CBAM pada Dataset Validasi

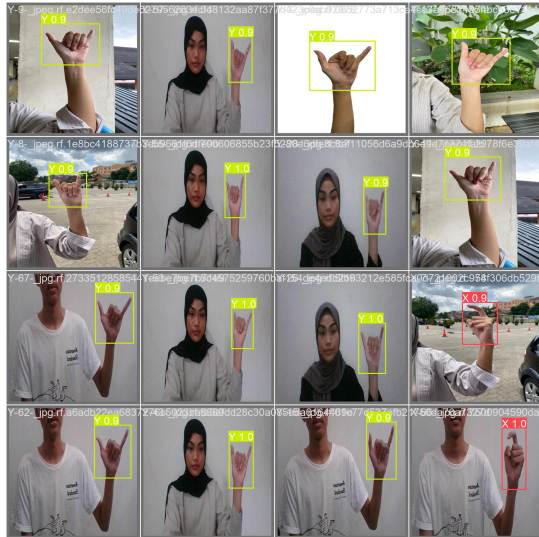
Warna diagonal utama yang dominan gelap menunjukkan tingkat akurasi tinggi (mendekati 1.0) pada mayoritas kelas. Kesalahan prediksi sangat minim, umumnya terbatas pada kelas dengan bentuk jari serupa seperti ‘M’ dan ‘N’.

2) *Peningkatan pada Kelas Sulit:* Berdasarkan data per kelas, model YOLOv8s-CBAM menunjukkan keunggulan spesifik:

- **Kelas “Jahat” (+2.1%):** Gestur dinamis ini melibatkan transisi bentuk jari. Peningkatan mAP menunjukkan CBAM membantu menangkap fitur temporal-spasial.
- **Kelas “Empat” (+1.4%):** Peningkatan signifikan membuktikan kemampuan modul *Spatial Attention* membedakan jumlah jari dengan presisi.
- **Kelas “Makan” dan “Bodoh”:** Model CBAM menunjukkan deteksi yang lebih robust pada gestur dengan

oklusi wajah, efektif memisahkan fitur tangan dari warna kulit wajah.

3) *Visualisasi Hasil Deteksi*: Gambar 3 memperlihatkan contoh hasil deteksi model pada data uji.



Gambar 3. Visualisasi Hasil Deteksi Model YOLOv8s-CBAM

Model mampu mendeteksi gestur SIBI dengan kotak pembatas (*bounding box*) yang rapat (*tight*) terhadap objek tangan dan label kelas yang akurat.

#### IV. KESIMPULAN DAN SARAN

##### A. Kesimpulan

Berdasarkan hasil implementasi, pelatihan, dan evaluasi yang telah dilakukan, penelitian ini menyimpulkan hal-hal sebagai berikut:

- 1) Integrasi modul atensi *Convolutional Block Attention Module* (CBAM) pada *backbone* arsitektur YOLOv8s berhasil diimplementasikan untuk deteksi Sistem Isyarat Bahasa Indonesia (SIBI), dengan tujuan memperkuat ekstraksi fitur pada gestur yang memiliki kemiripan visual tinggi.
- 2) Model YOLOv8s-CBAM yang dilatih dari awal (*scratch*) selama 100 epoch menunjukkan kapasitas pembelajaran yang kuat, mencapai performa yang sangat kompetitif (**mAP@0.5:0.95 sebesar 88.3%**) dibandingkan model *baseline* yang menggunakan bobot *pre-trained* COCO. Hal ini membuktikan bahwa arsitektur usulan mampu mempelajari representasi fitur SIBI secara efektif tanpa ketergantungan penuh pada transfer learning.
- 3) Mekanisme atensi terbukti efektif meningkatkan **Precision** deteksi menjadi **97.8%**, yang mengindikasikan penurunan tingkat kesalahan deteksi positif (*False Positives*). Penelitian ini membuktikan bahwa integrasi CBAM pada YOLOv8s mampu meningkatkan presisi deteksi gestur SIBI, terutama pada kelas-kelas yang mengalami oklusi wajah.

##### B. Saran

Berdasarkan keterbatasan penelitian ini, saran untuk pengembangan selanjutnya adalah:

- 1) Untuk penelitian selanjutnya, disarankan untuk mengintegrasikan *Transfer Learning* pada model ini untuk mempercepat konvergensi dan memperbaiki akurasi pada data yang lebih bervariasi.
- 2) Melakukan pengujian robustitas pada kondisi lingkungan yang lebih ekstrem dan bervariasi, seperti pencahayaan minim (*low-light*), latar belakang yang bergerak dinamis, atau sudut pengambilan gambar yang ekstrem, untuk menguji keandalan modul atensi di dunia nyata.
- 3) Mengeksplorasi variasi modul atensi lain atau kombinasi *pruning* model untuk mengembangkan versi yang lebih ringan (*lightweight*) agar dapat berjalan lebih efisien pada perangkat *mobile* dengan sumber daya terbatas.

#### PUSTAKA

- [1] S. Suharjito, R. Anderson, F. Wiryana, M. Ariesta, and G. Kusuma, "Sibi sign language recognition using convolutional neural network combined with transfer learning," *Pertanika Journal of Science & Technology*, vol. 29, no. 1, pp. 1–15, 2021.
- [2] M. Madyanto, R. Kurniawan, and Y. A. Wijaya, "Yolov8 algorithm to improve the sign language letter detection system model," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, no. 2, pp. 1379–1385, 2025.
- [3] W. Zhang, J. Li, and Y. Wang, "Improved yolov8 algorithm for small object detection in complex environments," *IEEE Access*, vol. 12, pp. 12 345–12 356, 2024.
- [4] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.