2025 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC) | 979-8-3315-2162-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICECCC65144.2025.11064241

Real-Time American Sign Language Detection Using YOLOv5 and YOLOv8: Analytical Comparison

Jeetu Rani
Assistant Professor
Department of Computer Science and Engineering
IFTM University, Moradabad, U.P, India
jeevanshi.chauhan@gmail.com

Bihari Nandan Pandey
Assistant Professor, Department of Computer Science and Engineering, Ajay Kumar Garg Engineering College, Ghaziabad, U.P, India
bnpanday@gmail.com

Mahesh Kumar Singh
Assistant Professor, Department of Electronics and Communication Engineering, Buddha Institute of Technology, Gorakhpur, U.P, India
mksingh39@bit.ac.in

Sachin Jain
Assistant Professor, Department of Computer Science and Engineering, Ajay Kumar Garg Engineering College, Ghaziabad, U.P, India
sachincs86@gmail.com

Prashant Upadhyay
Assistant Professor, Department of Computer Science, and Engineering, School of Engineering and Technology, Sharda University, Greater Noida, UP, Inida
prashanttheace@gmail.com

Sanjeev Bhardwaj
Assistant Professor
Department of Computer Science and Engineering
IFTM University, Moradabad, U.P, India
sanjeevmbd@gmail.com

***Abstract*- Culture and religion today are many and spread around the world. Since it began in 1817 at the American School for the Deaf (ASD), sign language culture has grown. Deep learning is used in computers today to solve problems and make apps that work in real time. Sign language (SL) is one of these. Convolutional neural network (CNN) is used in YOLO, an object detection and classification algorithm, to make it work quickly and correctly. The paper uses a custom model for sign language recognition to try to find American sign language using YOLO models and compare different YOLO algorithms. The tests showed that the new YOLOv8 did better than other versions of YOLO in terms of accuracy and mAP. However, YOLOv7 did better on the recall test than YOLOv8. The suggested model is quick and doesn't take up much space. It tests and trains on the American Sign Language letters dataset. It was able to achieve a 97 percent recall rate, a 95 percent precision rate, and a 96 percent mAP @0.5, which means it can recognise hand gestures in real time.**

Keywords- Convolution Neural Networks, Sign Language, YOLOv5, YOLOv8.

I. Introduction

What the WHO [1] says is that more than 1.5 billion people around the world have trouble hearing. Because they don't know how to use them right, about a billion teenagers are also at risk of losing their hearing. Children with these kinds of disabilities have a hard time with their physical and mental health, school, and job prospects. Older people often lose their hearing, become socially isolated, feel lonely, and get frustrated. Children who have trouble hearing may take longer to learn language and have trouble communicating. Unfortunately, people with hearing loss aren't always properly accommodated in public and private settings. This makes it harder for them to get jobs and do well in school. Children who have trouble hearing have a hard time understanding others and don't get much or any schooling. It is not possible for all deaf people to use the same sign language. The unique features of sign language and the difficulty of recording and understanding sign language make it hard to detect. Sign language is used in different ways in each country, even though there are some clear similarities.

Different languages are not recognised by the government or institutions, which is another big problem for the deaf community. The National Association of the Deaf [2] says that 18 million people in India have trouble hearing. Some new AI techniques [3] might be useful for an app that interprets sign language. Artificial intelligence (AI) is the study of making computers smart enough to learn from data they are given. These computers can decide what to do even when they don't know what's going on. Many systems and methods have been created to deal with these kinds of issues. This study is all about comparing two object detection algorithms and talking about what makes each one unique and what its strengths are [4]. Most schools favour American Sign Language (ASL) over Indian Sign Language (ISL). But ASL has been chosen in the project because it is suitable for model training with alphabets and numbers. The current research proposes a tailored Convolutional Neural Network (CNN) model with the capability of precise gesture prediction performed in real-time. The experimental functional model proves it is capable of recognizing hand gestures, hence beneficial in various real-world applications. “You Only Look Once” (YOLO) was created to implement a single-step procedure that could involve both classification and detection. YOLOv1 also referred to as YOLO detection network comprises 24 convolutional layers followed by maxpool layers and two dense layers are used at the end. Alternate convolutional layers are used to shrink the feature space between layers. The paper uses the pre-trained YOLO model on the Imagenet dataset for ASL detection. The GoogleNet [15] architecture is used as the base model for YOLO's architecture. The object detection task was implemented and evaluated on the VOC Pascal Dataset, and the model is trained on the Darknet framework. YOLO creates a likelihood of several bounding boxes for each grid cell. During training, one bounding box predictor must be in charge of each item.

II. Literature review

Sign language possesses a personality of its own since it is based on different types of body movement instead of particular sounds. Facial and gesture recognition has been extensively studied. Based on our findings, existing methods belong to two clear categories:

In the first approach, a separate external device is employed for sign recognition; whereas the second approach of hand gesture detection and recognition [10] is deep learning. Different researchers have proposed new SL recognition methods for cost, latency, performance, and portability. [5] describes how to train a classifier on a data set of 24 easy-to-spell gestures on the fingers using the "bag of visual words approach".

To turn each picture into a graph (a histogram) showing how often gestures were seen in that picture, this method first sorts the pictures into groups based on their features. The groups are then used to make a book with the gestures and how often they were seen. Two real-time hidden Markov model-based systems that only need a camera to watch the user help intermediate-level continuous ASL speakers [6]. The first device put the camera on the desk to watch the person using it, and it worked 92% of the time. The later device, on the other hand, puts the camera on the user's hat and is 98 percent accurate. [7] gave a detailed overview of recent progress in using deep learning to find objects in images. He talked about almost 300 different approaches, such as region-based object detection methods like SPPnet and Faster R-CNN, as well as classification and regression-based methods like YOLO [16]. In addition, the authors looked into and compared publicly available benchmark datasets, putting them into groups based on where they came from, what they were used for, their pros and cons, and the evaluation metrics that were used for each. With a Multi-layer Perceptron neural network trained on a dataset of 520 samples,[8] found a few signs. Table 1 shows a summary of the most up-to-date methods for finding sign language. For qualitative research, there are certain models that offer alternative frameworks or frameworks for gathering and interpreting information. Selecting a particular model can decide the study focus, the questions framed, the process of data collection, and its interpretation. Table 1 shows the qualitative analyses conducted with different models by different researchers.

Table 1. Compare new SOTA sign language detection methods.

Ref.	METHODOLOGY	DATASET	ALGORITHMS	ACCURACY Level
[2]	SL interpreter that uses image processing and machine learning	Six thousand pictures representing the 26 letters of the English alphabet.	HOG & SVM	88%
[5]	The bag-of-visual-words method for training SL classifiers	With a backdrop of plain white, 4972 images depict 24 static hand-pelling gestures	Automatic feature discovery uses SURF and BRISK. Evaluation uses KNN, SVM, and others.	SVM 91.35, KNN 86.97, BRISK Features 91.15, KNN 87.38
[8]	Leap Motion Controller real-time SL detection	There are 520 samples in 26 different ASL alphabets (consisting of 20 samples of each alphabet)	MLP, Backpropagation (A feature-based categorization model idea.)	96.15%
[9]	Leap Motion Sensor SL recognition	Four sets of data were gathered from the two signers, two sets from each person.	K-NN and SVM (A webcam and leap motion sensors could make learning in Second Life a lot better.)	KNN - 7.78%, SVM - 79.83%

Ref.	METHODOLOGY	DATASET	ALGORITHMS	ACCURACY Level
[11]	Live SL detection end-to-end.	5586 sign words and 26 alphabet signs	DeepSLR technology records coarse and finger movements with several sensors.	-
[12]	Real-time SL recognition at low cost	A, B, C, D, V signs from 10 users 50 signs A, B, C, D, V signs from 10 users 50 signs	Active shape models	76%
[13]	ML methods for ISL detection	800 single-handed Indian SL alphabet images and 220 double-handed ISL alphabet images	ISL	92%
[14]	Double-handed ISL dataset for testing machine learning classifiers	This is a dataset with 26 motions, and each one stands for an English letter.	HOG	87.67%

III. Proposed Methodology

Most of the time, Pascal VOC 2007 and Microsoft COCO [17] are used to find objects [20,21] and divide them into groups. This study uses YOLO versions 5 and 8 because they recognise sign language better. The suggested method is in Fig. 1.

3.1 YOLOv5

This test uses deep learning model YOLOv5 [18]. YOLOv5, made by Glenn Jocher, was the first model to be released without a scholarly paper. Its repository said that it was "ongoingly developing." Glenn Jocher works at Ultralytics LLC as a researcher. There is a GitHub page for YOLOv5 at https://github.com/ultralytics/yolov5/releases. The model came out in June 2020. Python [23], which is easier to install and integrate than earlier models written in C, was used to put YOLOv5 on Internet of Things (IoT) devices. The PyTorch community is also bigger than the Darknet community, which means it has more room to grow and develop in the future [19]. The speed is faster than other YOLO apps. CSPNET is what YOLOv5 uses as its main network for extracting feature maps. Path Aggregation Network (PANet) is also used to make the flow of information better. Figure 3 shows how YOLOv5 is put together. Here are some reasons why you should use YOLOv5:

1) The YOLOv5 contains SOTA aspects like an activation function, hyperparameter, data augmentation method and an easy-to-follow manual

2) The model is not complex in structure, thus easy to compute using minimal resources.

3) YOLOv5's compactness and lightness allow it to be used in mobile phones and embedded systems.

3.2 YOLOv8

Besides that, YOLOv8 is the newest model in the series of algorithms. Glenn Jocher wrote and posted it January 23, 2023. Anchor-free detection and mosaic augmentation are new features of the unfinished version. Model training is easier with YOLOv8's CLI. A Python package simplifies development compared to the old model. This is where you can find the YOLOv8 repository on Github: https://github.com/ultralytics/ultralytics. Soon after the image is processed just once, predictions are made for both the bounding boxes and the classes. The algorithm works because it does both predictions at the same time: the bounding box and the classification. Confidence is the probability of an object being in each bounding box[22]. Both YOLOv5 and YOLOv8 algorithms must be installed first. Both algorithms will run simultaneously on devices with the same specs, making comparisons easier and saving time. The Roboflow public data set "American Sign Language letters" [23] is used in this project. PyTorch, based on Torch, will also be used. The Python-based computer vision and natural language processing library. A number of ways are given for how to use a downloaded data set in the model when it is downloaded from Roboflow. PyTorch can set up a roboflow package and download the data set to the programme directory. Ultralytics is another large download. This package has all YOLO algorithm versions, making it ideal for this programme.

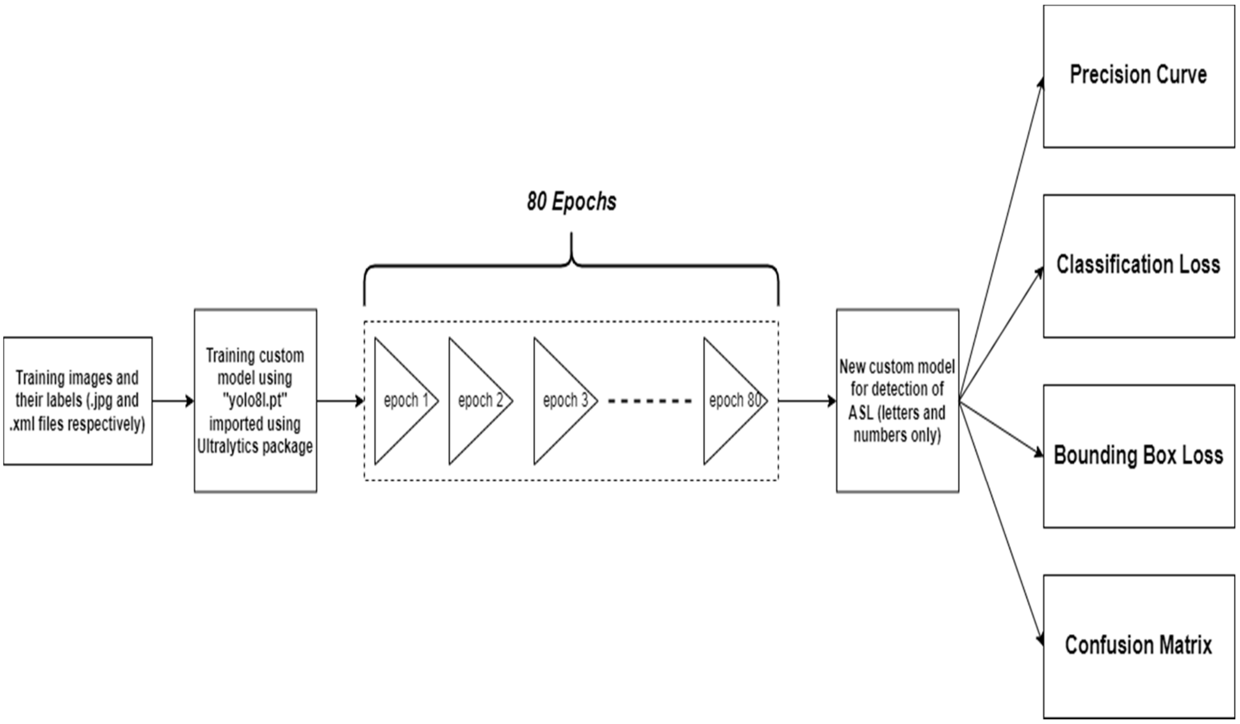
Training images and their labels (.jpg and .xml files respectively)
Training custom model using "yolo8l.pt" imported using Ultralytics package
80 Epochs
epoch 1
epoch 2
epoch 3
epoch 80
New custom model for detection of ASL (letters and numbers only)
Precision Curve
Classification Loss
Bounding Box Loss
Confusion Matrix

Fig. 1. Steps to summarize the video

IV. Comparative analysis

The dataset has 72 test, 1512 training, and 144 validation images, a 1:21:2 ratio. New models are trained for 80 epochs. ASL dataset confusion matrix for YOLOv8 and v5 is shown in Fig. 2. YOLOv8's confusion matrix is more varied than v5. This makes YOLOv8 less accurate for new images and real-world situations. Except for the background, YOLOv5 predicts different values twelve times. However, YOLOv8 has 13 such events. This comparison may seem pointless, but the two matrices show that the background predictions for both cases are very different. Iterations to reach the minimum vary between v5 and v8. V8 requires fewer iterations than V5. Maybe because YOLOv8 is still in development compared to its predecessor.

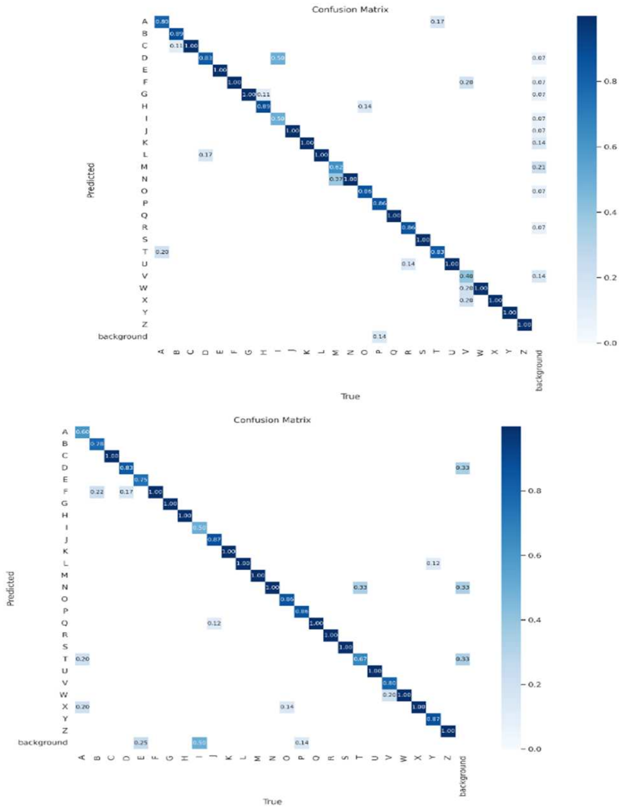

Fig. 2. Confusion Matrix for YOLOv5 and YOLOv8

The other model has 7 cases of wrong predictions and their importance. Fig. 3 compares YOLO loss reduction, both versions. V5 decreased faster than v8. Both bounding box and classification loss minimum points are faster in YOLOv8 around 40 epochs. In contrast, Table 2 shows that YOLOv5 results point after the 50th epoch, likely around the 60th or higher. The custom model correctly detected hand gestures 96% of the time and in real time 96%. In the v5 and v8 base models, bounding boxes lose 0.326 and 0.312 and classification loses 0.191 and 0.175. This shows YOLOv8 finds and classifies better. Object detection and hand gesture reading are faster in v8. Early YOLOv5 points are crowded and high. After training, models were tested in a small validation set. The surface results show that both models work well when the gesture is visible to the naked eye. Uncertain gestures that are mistaken for letters are different. This app answers better. It predicts unspoken gestures. YOLOv5 predicted well.

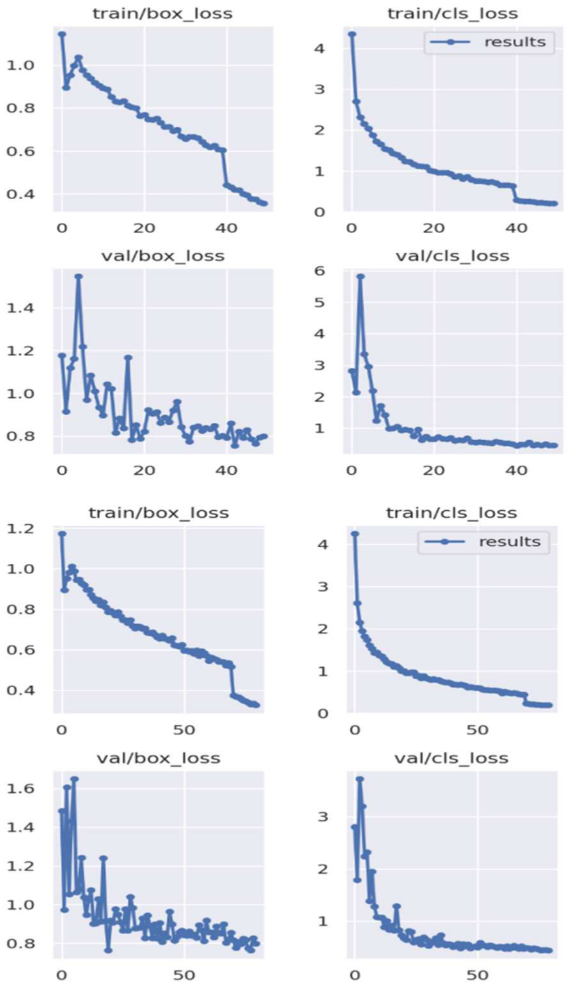

Fig. 3. Loss reduction for YOLOv5 (First) and YOLOv8 (Second)

Table 2. Loss for YOLOv5 and YOLOv8

Loss and Epoch		*YOLOv5*	*YOLOv8*
"BOUNDING BOX LOSS"	1st epoch	1.733	1.444
	80th epoch	0.3265	0.312
"CLASSIFICATION LOSS"	1st epoch	4.251	4.352
	80th epoch	0.191	0.1735
"mAP"		93.60%	96%

V. Conclusion

Training, validation, and testing are better with YOLOv8. YOLOv8 is still a theory because the model is not ready for practical use in ongoing movement. In comparison, YOLOv5 can predict the correct gesture easily and even completes the blanks that were left by the other model. This is an example of how handy our system can be, resulting in greater satisfaction and usability. In the future, further advancements in NLP and machine learning can make video summarization systems even more powerful, making content navigation easier and more effective in the digital age.

References

[1] Johri M, Téhinian S, Pérez Osorio MC, Barış E, Wahl B. Vaccination for prevention of hearing loss: a scoping review. Communications Medicine. 2025 Mar 24;5(1):85.

[2] Dixit A, Jain S, Kazmi A, Manorama K. Gesture Based Model using Deep Learning. In2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET) 2024 Aug 23 (pp. 1-5). IEEE.

[3] Halvardsson, Gustaf, Johanna Peterson, César Soto-Valero, and Benoit Baudry. "Interpretation of swedish sign language using convolutional neural networks and transfer learning." SN Computer Science 2, no. 3 (2021): 207.

[4] Trivedi NK, Jain S, Kaswan S, Jain V. Federated Learning Empowered Breast Cancer Detection in Images: A YOLO and ResNet-50 Fusion Approach. In2024 International Conference on Computational Intelligence and Computing Applications (ICCICA) 2024 May 23 (Vol. 1, pp. 24-29). IEEE.

[5] Jain S, Jaidka P. Lung Cancer Classification Using Deep Learning Hybrid Model. InFuture of AI in Medical Imaging 2024 (pp. 207-223). IGI Global.

[6] Starner T, Weaver J, Pentland A. Real-time american sign language recognition using desk and wearable computer based video. IEEE Transactions on pattern analysis and machine intelligence. 2002 Aug 6;20(12):1371-5.

[7] Aziz L, Salam MS, Sheikh UU, Ayub S. Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review. Ieee Access. 2020 Sep 3;8:170461-95.

[8] Naglot D, Kulkarni M. Real time sign language recognition using the leap motion controller. In2016 international conference on inventive computation technologies (ICICT) 2016 Aug 26 (Vol. 3, pp. 1-5). IEEE.

[9] Chuan CH, Regina E, Guardino C. American sign language recognition using leap motion sensor. In2014 13th International Conference on Machine Learning and Applications 2014 Dec 3 (pp. 541-544). IEEE.

[10] Jain S, Jain V, Chatterjee JM. Ensemble based brain tumor classification technique from MRI based on K fold validation approach. Journal of Integrated Science and Technology. 2025 Mar 14;13(5):1114-..

[11] Wang Z, Zhao T, Ma J, Chen H, Liu K, Shao H, Wang Q, Ren J. Hear sign language: A real-time end-to-end sign language recognition system. IEEE Transactions on Mobile Computing. 2020 Nov 16;21(7):2398-410.

[12] Fernando M, Wijayanayaka J. Low cost approach for real time sign language recognition. In2013 IEEE 8th International Conference on Industrial and Information Systems 2013 Dec 17 (pp. 637-642). IEEE.

[13] Dutta KK, Bellary SA. Machine learning techniques for Indian sign language recognition. In2017 international conference on current trends in computer, electrical, electronics and communication (CTCEEC) 2017 Sep 8 (pp. 333-336). IEEE.

[14] Jain S, Jain V. Ensemble Techniques for Classification of Brain Tumor Images Based on Weighting Average of Various Deep Learning-Based Components Models. International Journal of Performability Engineering. 2023 Oct 28;19(10):676.

[15] Trivedi NK, Jain S, Misra A, Tiwari RG, Maheshwari S, Gautam V. Skill-Honey Badger Optimisation Algorithm-Enabled Deep Convolutional Neural Network for Multiclass Leaf Disease Detection in Tomato Plant. Journal of Phytopathology. 2024 Nov;172(6):e70001.

[16] Kyranou I, Szymaniak K, Nazarpour K. EMG dataset for gesture recognition with arm translation. Scientific Data. 2025 Jan 17;12(1):100.

[17] Prakash KS, Kunju N. An optimized electrode configuration for wrist wearable EMG-based hand gesture recognition using machine learning. Expert Systems with Applications. 2025 May 15;274:127040.

[18] Le VH. Selected hand gesture recognition model based on cross-evaluation of deep learning from large RGB image datasets. Multimedia Tools and Applications. 2025 Mar 20:1-50.

[19] Nguyen TH, Ngo BV, Nguyen TN. Vision-Based Hand Gesture Recognition Using a YOLOv8n Model for the Navigation of a Smart Wheelchair. Electronics. 2025 Feb 13;14(4):734.

[20] Jain S, Jain V. Novel hybrid boosted ensemble learning framework for brain tumor prediction. In2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) 2022 Mar 23 (pp. 866-869). IEEE.

[21] Diwan, Tausif, G. Anirudh, and Jitendra V. Tembhurne. "Object detection using YOLO: Challenges, architectural successors, datasets and applications." multimedia Tools and Applications 82, no. 6 (2023): 9243-9275.

[22] Trivedi NK, Maheswari S, Sharma H, Jain S, Agarwal S. Early Detection & Prediction of Heart Disease using Various Machine Learning Approaches. In2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) 2022 Mar 23 (pp. 793-797). IEEE.

[23] Lee, David. "American Sign Language Letters Dataset." Public Domain (2020).