Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Comparative Analysis of Various Yolo Models for Sign Language Recognition with a specific dataset

Mangai .V[1], Dr. Kalaimagal .R[2]
[1]Research Scholar, PG & Research Department of Computer Science, Government Arts College. Nandanam, Chennai – 35
[2]Associate Professor, PG & Research Department of Computer Science, Government Arts College. Nandanam, Chennai – 35

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Understanding and replaying sign language are the hardcore communication tasks between the normal person and to deaf and dumped person, and vice versa. To enhance sign language-based communication, several models have been developed for making sign language into an understandable format by translating gestures into words. The ultimate goal of this research paper is to analyse and compare the various You Only Look One (YOLO) models on SLR problem. YOLO is a fast and efficient convolutional neural networks (CNN) variant that provides a better solution for sign language problems. The comparison of different YOLO models with Indian Sign Language (ISL) dataset can provide a suitable YOLO model for SLR. Therefore, the proposed work has considered the ISign Benchmark dataset. The ISL-based comparison analysis is implemented on Python tool where the various performance metrics are calculated for selecting the best YOLO model . This will make a way to give a fast and efficient means for recognizing sign gestures.

Keywords: SLR, CNN, YOLO model, Dataset, Performance metrics

I. INTRODUCTION

Sign Language is the main part of communication with the people of hard-hearing. This kind of people when they want to be included in the society, problem arises them to communicate with other common people. It creates a barrier between the two. Also they require a sign interpreter to move within the society. Even normal people find it difficult to understand the sign without the explanation of the interpreter. To come out from this barrier, there should be a system to bridge the gap between the public and muted community. Despite the growing awareness and advocacy for inclusivity, technological solutions that facilitate seamless communication between signers and non-signers remain limited, particularly in the Indian context where ISL (Indian Sign Language) lacks mainstream technological support. In recent years, advances in computer vision and deep learning have enabled the development of Sign Language Recognition (SLR) systems capable of translating gestures into spoken or written language. These systems have the potential to enhance accessibility in various settings such as education, healthcare, and public services. Among the various sign languages studied globally, ISL presents unique challenges due to its complex grammar, regional variations, and limited publicly available datasets. Furthermore, the majority of SLR research has focused on American Sign Language (ASL), leaving ISL underrepresented in terms of both resources and research contributions.

Early ISL recognition systems relied on sensor-based gloves or marker-based vision systems, which, although effective, were neither practical nor scalable for everyday use. The emergence of deep learning has shifted the focus toward visionbased solutions, with Convolutional Neural Networks (CNNs) being employed to classify static hand gestures. Object detection algorithms—especially YOLO (You Only Look Once)—have gained popularity due to their ability to detect and classify objects in a single pass, offering both speed and accuracy. YOLO-based models have shown significant promise in SLR by enabling real-time gesture recognition with efficient hand localization. These properties make YOLO a strong candidate for developing scalable, real-time ISL recognition systems that can function reliably in real-world scenarios.

This research aims to explore the application of YOLO for Indian Sign Language recognition, focusing on its effectiveness in recognizing isolated hand gestures in real time. The study also investigates the challenges involved in building robust ISL recognition systems, including dataset limitations, signer variability, and environmental noise. By leveraging the strengths of YOLO, the proposed system seeks to contribute toward accessible and deployable solutions for the ISL community.

Literature Review

Ashish Sharma et al.,[3] focussed on hand gesture recognition using Image Processing and Feature extraction techniques. The model is constructed to recognize sign gesture images of the hand, which utilizes Oriented FAST and rotated BRIEF as a feature detector, having efficacy and better performance better than widely used feature detectors such as SIFT and SURF,etc.,.The model utilizes a combination of several other techniques

and classifiers.Ayushi N.Patani et al.,[4] aimed to acquaint a sign language communication works and puts forward research conducted in this field that explains how to capture and recognize sign language and also attempts to suggest a systemized solution. This model provides a simple, seamless and highly available means to communicate with other members of society

Suharjito et al.,[1] gave a review based on Sign language recognition application system. They gave a clear study of data acquisition methods from different research anders and their classification algorithms compared with its accuracy. Also they discussed about the limitations in their reviewed papers.

Sanket Bankar et al.,[6] designed a YOLOv5 model to detect gesture recognition. The model focusses on labelled dataset Roboflow. They compared their model with CNN. Their dataset consists of hand images depicted hand signs having different angles and backgrounds. They recognized alphabets successfully.

Rachana Patil et al.,[5] their study used web camera for capturing hand gestures. The captured image undergoes series of processing steps which include various computer vision technique such as conversion to gray-scale, dilation and mask operation. They applied CNN method to train and identify pictures. They used ASL dataset.

Refat Khan Pathan et al.,[8] used ASL dataset with alphabets and numbers They recognize using the fusion of image and hand landmarks through multi-headed convolutional neural network. The model applied a fusion of traditional image with extracted hand landmarks and trained on a multiheaded CNN.

Pallavi P et al.,[10]given a model for recognizing sign language using deep neural network. They designed a system to perform accurately using YOLOV3.The method uses ASL dataset for training. In the trained model average loss is calculated, Here IOU and mAP are considered as performance measure. The prediction percentage and bounding boxes accuracy in this model depends on Batch sixze, Learning rate, and number of training iterations.

Sreemathy. R et al.,[9]The study proposes a python based system that classifies words from sign language. In this work they proposed two models YOLOV4 and SVM with media pipe. The trained models were extended to be able to recognize continuous gestures with high accuracy with Gramformer. The model focussed to give better accuracy when compared with related models.

There are many Challenges in detecting the sign gesture from the available dataset. Communication is fundamental for social participation. The main motivation is to break the gap between the deaf and dumb people and the general public. Communication is fundamental for societal participation. Individuals who are deaf and mute utilize sign language to interact, yet those without hearing impairments often struggle to comprehend this mode of communication. While

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

substantial research has been conducted on American Sign Language (ASL), Indian Sign Language (ISL) presents distinct differences. Notably, ISL employs both hands for signing, whereas ASL typically uses one. The use of both hands can lead to feature ambiguity due to hand overlap. Moreover, the scarcity of datasets and regional variations in ISL have limited progress in gesture recognition. The study seeks to bridge the communication gap between hearing individuals and those who are deaf and mute by focusing on ISL. Expanding this initiative to include words and common expressions could facilitate more efficient and effortless communication for the deaf and mute community, while also advancing the development of autonomous systems to assist them.

.With the advent of deep learning, Convolutional Neural Networks (CNNs) and real-time object detection algorithms such as YOLO have significantly advanced the state-of-the-art in this domain. Future research can focus on hybrid models that combine YOLO for spatial detection with sequence learning algorithms for temporal recognition. Additionally, efforts to create standardized, multilingual ISL datasets with signer variation and real-world noise will be essential for scaling SLR systems for widespread use in India.

I. PROPOSED SYSTEM

Finding out an efficient model to accurately perform the task of hand gesture recognition using deep learning technique, YOLO algorithms. The main focus in designing the model is to handle occlusion.

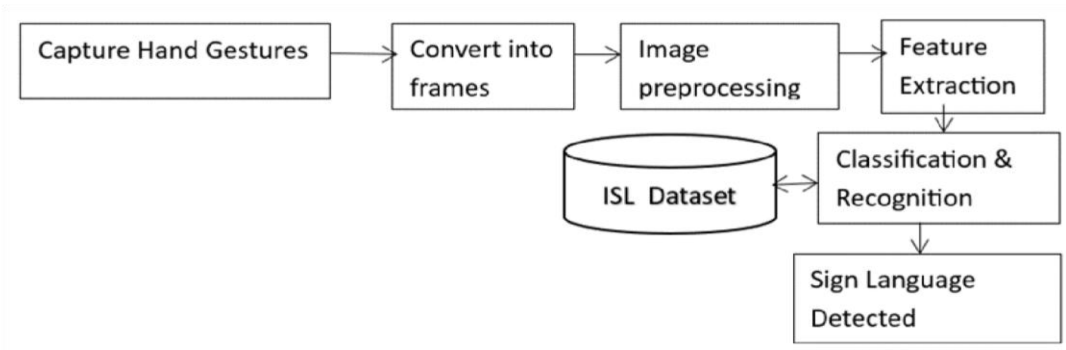

Fig.1. Block diagram of the system

The proposed system focuses on recognizing Indian Sign Language (ISL) gestures using various versions of the YOLO algorithm, trained and evaluated on the iSign benchmark dataset. The system is designed for real-time hand gesture detection, targeting robust performance in diverse lighting, background, and signer conditions. The architecture is modular, allowing plug-and-play use of different YOLO models (YOLOv3, v4, v5, v7, or v8) for benchmarking and optimization. The YOLO (You Only Look Once) algorithm offers a unified approach to object detection, integrating feature extraction and object localization into a single, efficient process. This architecture is particularly advantageous for realtime applications, such as Sign Language Recognition (SLR), where rapid and accurate detection of hand gestures is essential.

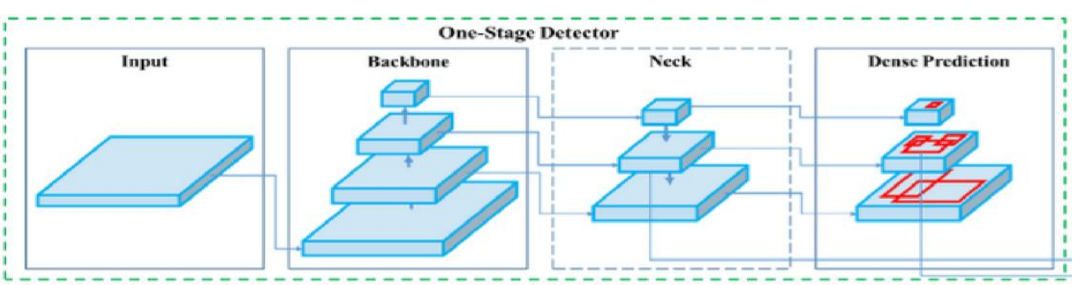

Fig. 2. Simple Yolo Architecture

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

The Key component includes Input Module, which Supports both real-time camera input and pre-recorded ISL video/image sequences from the iSign dataset. Preprocessing, enhances the image quality and ensures compatibility with YOLO input. YOLO Detection Module which evaluates multiple YOLO versions for performance comparison. Detects and localizes one or both hands in each frame and Outputs bounding boxes with class confidence scores. Classification Layer, assigns gesture labels to detected hands based on YOLO's predictions. In custom-trained YOLO models, gesture classes are pre-defined during training on the iSign dataset. Output Interface, displays the recognized gesture as text

II. METHODOLOGY

This section outlines the step-by-step process employed to develop and evaluate an Indian Sign Language (ISL) recognition system using the ISign dataset and various versions of the You Only Look Once (YOLO) object detection algorithm. The methodology comprises data preparation, model selection, training, evaluation, and comparative analysis.

Operational Workflow of YOLO algorithms in SLR:

Input Processing: The input image is divided into an S×S grid. Each grid cell is tasked with predicting bounding boxes and confidence scores, indicating the presence and accuracy of an object within the cell.

Feature Extraction: The backbone processes the image to extract hierarchical features, capturing both low-level details and high-level semantic information pertinent to hand gestures.

Feature Aggregation: The neck combines these features, ensuring that contextual information is preserved and that the model remains sensitive to variations in hand shapes and positions.

Prediction: The head generates bounding box coordinates and class probabilities for each grid cell, identifying the location and type of gesture present.

Post-Processing: Techniques such as Non-Maximum Suppression (NMS) are applied to eliminate redundant detections, refining the final set of recognized gestures.

By integrating these components, YOLO streamlines the detection process, offering a robust solution for real-time Sign Language Recognition. Its architecture ensures that both the spatial and semantic nuances of hand gestures are effectively captured, facilitating accurate and efficient interpretation.

Each YOLO model's performance was evaluated based on:

Accuracy is calculated for finding the overall correctness of the model

$$\text{Accuracy} = (TP+TN)/(TP + TN + FP + FN)$$

Precision is evaluated for finding how many predicted positives are actually correct

$$\text{Precision} = TP / (TP + FP)$$

Geometric mean (G- Mean) to evaluate the performance across both classes by rewarding balance

$$\text{G-Mean} = \sqrt{(sensitivity\ x\ specificity)}$$

Pearson Correlative Co-efficient (PCC - r) to find the measures linear correlation between predicted labels and true labels

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \cdot \sqrt{\sum (y_i - \bar{y})^2}}$$

Finally the obtained result is analysed and the best YOLO model has been figured out.

Results

The dataset iSign is taken for training the various versions of YOLO model. The dataset contains gestures of numbers and alphabets. Sample images from the dataset were shown in the fig. 3.

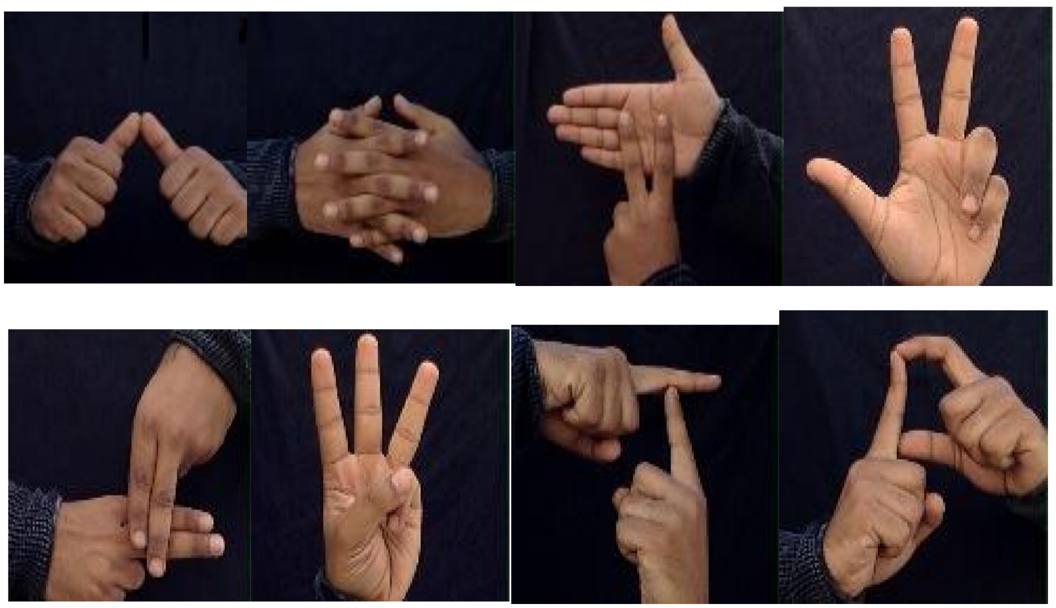

Fig. 3. Few images from iSign dataset

The selected YOLO model has been configured and training process of the same were completed. The fig.4, shows some of the predicted images.

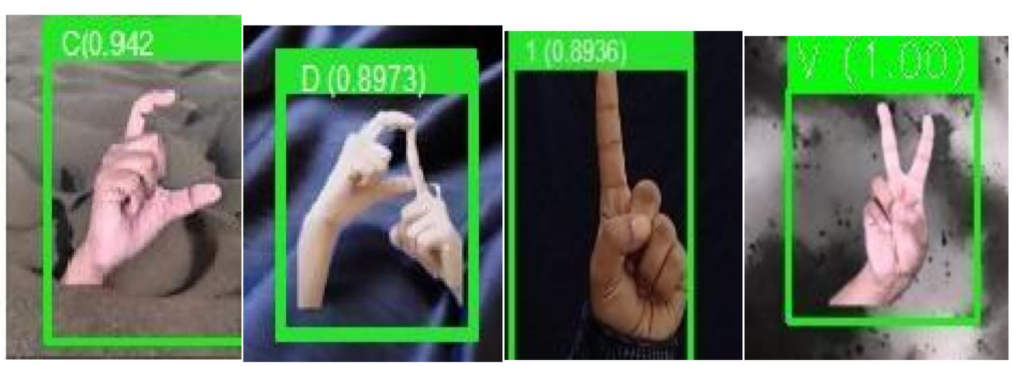

Fig. 4. Few of the predicted images

The below table (Table No. 1) shows the performance result of isign dataset in various YOLO versions

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

Table I. Performance

ISign Benchmark dataset	YOLOv4	YOLOv5	YOLOv6	YOLOv7	YOLOv8
Accuracy	0.90	0.92	0.94	0.95	0.96
Precision	0.90	0.91	0.93	0.95	0.96
Balanced accuracy	0.89	0.92	0.93	0.94	0.95
Geometric mean	0.90	0.92	0.94	0.96	0.97
Mathew's correlative coefficient	0.88	0.91	0.93	0.95	0.96
Pearson's correlative coefficient	0.90	0.91	0.93	0.95	0.96
Computation Time	0.02	0.02	0.02	0.02	0.02

The table shows a comparative result of YOLO versions. Except version 4 remaining versions show an accuracy with more than 90 % . Remaining all metrics also shows the same performance. YOLOv8 gives a best value among the taken versions of YOLO.Thus, from the tabulation, we realize the calculation of each high version shows an improvement than its previous versions. Among the versions of YOLO compared YOLOv8 gives a better result than others.

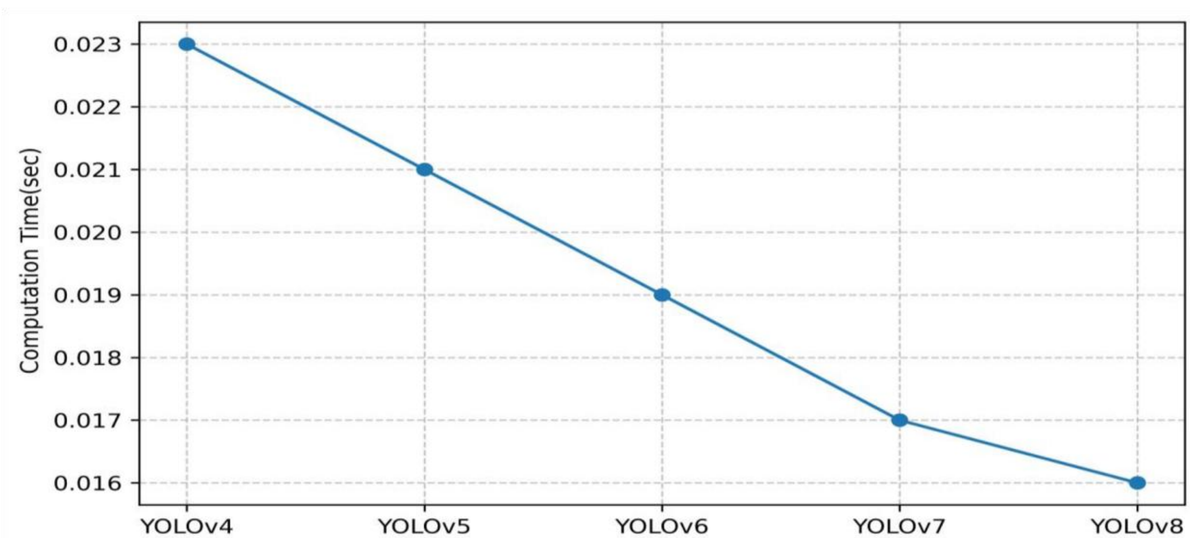

Fig. 5. Computation Time Chart

The chart (Fig.5) gives the clear picture of the Computation time of the algorithms used and the time taken to process with the iSign dataset. Here also it shows how time decreases on the higher versions.

III. CONCLUSION

The study focuses on the study of the YOLO v4, YOLOv5, YOLOv6, YOLOv7 and YOLOv8 algorithms for recognizing hand gestures with iSign dataset. Accuracy and Precision show overall recognition and confidence in detected signs. Balance Accuracy and Geometic mean addresses the class imbalance in sign types. Mathew correlative co-efficient gives a single robust metric accounting of all outcomes. Pearson correlative co-efficient which can be used to access similarity in confidence scores. In future, the higher versions with real time dataset can be followed to give a better accuracy.

Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

REFERENCES

[1] Suharjito, Ricki Anderson, Fanny Wiryana, Meita Chandra Arista, Gede Putra Kusuma(2017) Sign Language Recognition Application System For Deaf -Mute People: A Review Based on Input-Process-Output, In: Elsevier B.V

[2] E. Padmalatha, S, Sasilekya, R. Ravinder Reddy, Ch. Anil Krishnan, K. Divyarsha(2019) Sign Language Recognition, In: IJRTE.

[3] Ashish Sharma, Anmol Mittal, Savitoj Singh, Vasudev Awatramani(2020) Hand Gesture Recognition using Image Processing an Feature Extraction Techniques. In:International Conference on Smart sustainable intelligent computing and applications under ICITETM, Elsevier.

[4] Ayushi N. Patani, Varun S. Gawande, Jash V. Gujarathi, Vedant K. Puranik, Tushar A. Rane(2021) Methodologies for Sign Language Recognition: A Survey. In:International Journal of Innovative Science and Research Technology.

[5] Ratchana Patil, Vivek Patil, Abhishek Bahuguna and Mr. Gaurav Datkhile(2021) Indian Sign Language Recognition using Convolutional Neural Network, In: ICACC.

[6] Sanket Bankar, Tushar Kadam, Vedant Korhale, Mrs. A. Kulkarnai(2022) Real Time Sign Language Recognition Using Deep Learning, In: IRJET..

[7] Vijeeta Patil, Sujata C, Shridhar Allagi, Balachandra Chikkoppa(2022) A Deep Learning Framework for Real-Time Sign Language Recognition Based on Transfer Learning, In: IJATT.

[8] Refat Khan Pathan, Munmun Biswas, Suraiya Yasman, Mayeen Uddin Khandaker,Mohammad Salman, Ahmed A.F.Youssef(2023) Sign Language Recognition using the fusion of image and landmarks through multi-headed convolutional neural networks, In: Scientific Reports.

[9] R Sreemathy, MP Turuk, S Chaudhary, K Lavate, A Ushire, S Khurana(2023) Continuous word level sign language recognition using an expert system based on machine learning, In: International Journal of Cognitive Computing in Engineering.

[10] Pallavi P, Saravamangala D.R(2021) Recognition of Sign Language using Deep Neural Network

[11] Chiranjeevulu D, Tejasri J, Sasikala I, Hemanth K, Rakesh N (2023) Real Time Sign Language Detection using YOLOv5 algorithm in IJRPR