# Real Time Recognition of Sign Language Using Convolutional Neural Network

SEJAL VASAN[1], TUSHAR BHUTANI[2], AMITA GOEL[3], NIDHI SENGAR[4], VASUDHA BAHL[5]
[1, 2, 3, 4] *Maharaja Agrasen Institute of Technology, Delhi, India*

***Abstract-*** ***Around 1 million to 2.7 million Indians use Indian Sign Language to communicate; this figure clearly states how important it is to augment a strong and dependable system for easy communication. Sign language is mostly learned by the deaf and dumb and is most likely unknown to others; as a result, communication becomes difficult. With the advancement of time, various approaches have emerged for smooth communication using sign language. The most common technique for interpretation involves using image processing algorithms to draw out features from coordinated motions, then applying convolutional neural networks (CNN) to master these characteristics and improve their functionality. An algorithm that can identify and predict objects in one forward pass is You Only Look Once (YOLO). In this study specifically Yolov7 i.e., the latest model of the Yolo series was used. The development of a system that enables individuals to utilise sign language independently can significantly aid them in being independent and ignite the confidence to exhibit their individuality to the public fearlessly. Therefore, it is crucial to create a hand gesture recognition system that can recognise hand signs in a developing nation like India with utmost accuracy. Our study dealt with developing a static hand gesture recognition algorithm to detect Indian Sign Language gestures used to communicate in our day-to-day lives. The machine learning model built was able to provide a perfect precision-recall graph, with 99% precision and recall.***

***Indexed Terms-*** ***Indian Sign Language, Convolutional Neural Networks, You Only Look Once***

## I. INTRODUCTION

### *A. Indian Sign Language*

Sign languages are visual languages that predominantly use facial and hand movements. There are 100 or more different sign languages around the world, and the motions or symbols of these are arranged linguistically.

Indian Sign Language (ISL) is the primary form of conversation among deaf people in India. Inan interview for The Wire, Anuj Jain (executive director of the National Association of the Deaf) said that over 98% of hearing-impaired youngsters drop out of school or stay illiterate in the absence of teachers using sign language. He suggested that, for educational purposes and to help the deaf find employment in the private and public sectors, standardising Indian Sign Language (ISL) is crucial. Some techniques that have been used previously in order to build real time systems for sign language recognition, some of them primarily use skin segmentation features of Open CV [1], svm [2] and Convolutional Neural Network [3]. In this study, the dataset created included a few Indian gestures as per ISL, however most gestures in Indian Sign Language are not static but a combination of two or more gestures. This study was restricted to the gestures that can be shown using a single static gesture.

### *B. Convolutional Neural Networks*

The most leading and appropriate technology that till date have been used for hand motion identification is CNN.

Because it contains the majority of the computing, the convolutional layer is the most important component of a CNN. Its three parts are input data, a filter, and a feature map. Numerous CNN architectures have been presented in recent years [4]. The architecture of convolutional neural networks consists of several

convolutional layers. Because of how closely this architecture resembles how neurons are arranged in the human brain, the term "Neural Network" was created.

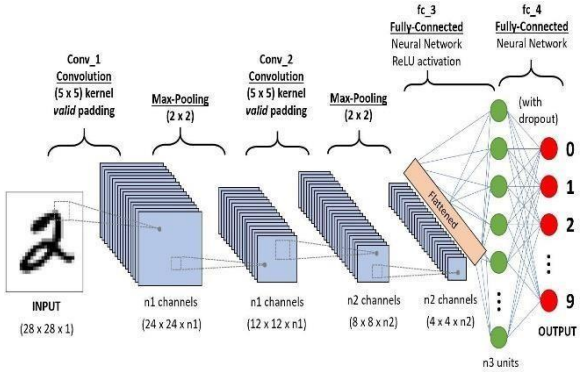


Fig. 1. CNN Architecture

According to how the CNN algorithm works, a picture is utilised as an input, quality-based weights and biases are applied, and the image is then used as a dataset in accordance with project requirements.

*C. You Only Look Once*

A fast multi-entity identification technique called YOLO (You Only Look Once) was first described in a paper in a 2015 [5] study from the University of Washington. Since then, other enhancements have been suggested.

YOLO v7 was the version that we employed in our research. All previous object identification algorithms and YOLO iterations are outperformed in terms of speed and accuracy by the YOLO v7 [6] method. It requires technology that is several times less expensive than other neural networks and can be trained substantially faster on small datasets without any pre-learned weights. As a result, it is projected that YOLOv7 will soon surpass YOLOv4, which was the previous state-of-the-art for real-time applications, and take over as the accepted industry standard for object detection.

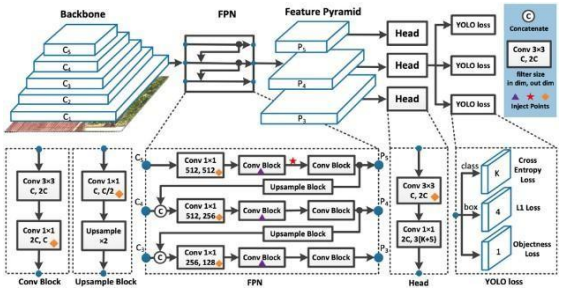


Fig. 2. YOLO network architecture as shown in PP-YOLO[7]. PaddlePaddle is a deep learning framework written by Baidu.

The model trained in our study was trained using a dataset that included multiple gestures captured using a Python script. Further, we used YOLO v7 algorithms to check accuracy and precision. After configuring YOLO v7 paths and data annotation, followed by downloading, configuring pre-trained YOLO v7, and training and testing of the model were done.

## II. RELATED WORK

This study is a continuation of the work from [8]. In this research, we have discussed the most recent repetition, which incorporates significant advancements in You Only Look Once versions over the previous years. We have also looked at other studies that discuss sign language interpreters. It is vital to present an overview of research on both languages, that is both Indian Sign Language and American Sign Language. This is because there is vast study on American Sign Language which provides a good incentive to go through its previous work.

Using dynamic hand gesture recognition algorithms [9], the authors made an effort to recognise ISL real time movements. To capture motion, they used high-resolution videos. For pre-processing, the video was first transformed to the HSV colour space, and then skin pixels were used to segment gestures. The dataset was produced using the Support Vector Machine (SVM), which was then used to recognise the movements.

Indian Sign Language gestures have been attempted to be categorized [10] utilising a classification approach based on Euclidean distance. According to

this study article, their method consists of the following steps: Feature extraction, feature categorization, hand cropping, and skin filtering.

Deora, Divya, and Bajaj, Nikesh [11] In their article, they outline a framework for a human- computer interface that can translate gestures in Indian Sign Language (ISL). Its key finding is that gesture detection becomes more challenging as both hands are employed. Therefore, separation becomes difficult.
According to the study cited [12] they tried developing a programme that can instantly recognise movements made in American Sign Language using YOLO (You Only Look Once) method. The programme begins with data collection, goes on to pre process gestures, and then uses a combinatorial algorithm to trace hand movement.

Different Indian Sign Language gestures have been captured in a video dataset [13]. Deep learning and support vector machines were both used to categorize the gestures. People between the ages of 22 and 26's hand motions were employed to obtain the data. This has made it easier to create a database, which is essential for any research project.

Using ISL data that was already available, the authors [14] developed a dataset. They used direction holograms for categorization because of their appeal for lighting and orientation invariance. KNN and the Euclidean distance were also applied.

An online application was developed by Bui, Hien Minh to assess the validity of applying the Yolo model to translate the alphabets of American sign language [15]. The findings of this thesis project demonstrate that Yolo version 3 was not the ideal option for sign language, despite its remarkable speeds and efficiency in static object identification. So, while doing our study we had to use the Yolo version, which is more dependable.

In [16] the authors proposed a study on the use of deep learning techniques to recognize hand signs in the context of hand motions and greetings from Indian Sign Language. The authors used a dataset of Indian hand signs and gestures and evaluated the performance of various deep learning models on this dataset.

The authors [17] reviewed the state of text-to- Indian sign language translation systems in their comprehensive review. They examined various methods and approaches used in existing systems, and evaluate their performance. They also discuss the challenges and future directions for research in this area.

Using a combination of ensemble-based classifiers, Indian sign language was studied and recognised [18]. To increase the accuracy of Indian sign language recognition, they suggested combining classifiers. On a dataset of motions in Indian sign language, they assessed the performance of their suggested strategy and contrasted it with other ones which already exist.

This paper [19] presented an approach for recognizing hand gestures using deep learning. It suggested employing a depth sensor to record hand motions, and then a convolutional neural network (CNN) to identify the gestures. According to the authors, the proposed approach on a dataset of 25 different hand gestures has an accuracy rate of 96.67%.

A novel dataset for the real-time hand gesture detection of Bangla sign digits was reported in this study [20]. The paper introduced a dataset of 10 Bangla sign digits, which were captured using a depth sensor and included both left- and right-hand gestures. The dataset, according to the authors, can be utilised to create systems for the deaf and hard of hearing people that can recognise hand gestures for Bangla sign digits in real time.

A method [21] for detecting accurate dynamic hand gestures based on sign words using CNN with feature fusion. The authors propose to use a CNN to extract features from a sequence of depth maps, and then use feature fusion to combine the features from different frames in the sequence. They evaluated their proposed method on a dataset of dynamic hand gestures and reported that it achieved an accuracy of 96.72%, which is a significant improvement over other methods that have been proposed in the literature.

You Only Look Once (YOLO) network-based object detection technology was demonstrated in this study [22]. The authors proposed to use the YOLO network to detect objects in images, and evaluate their proposed method on a set of pictures. They claimed that the suggested approach had a high degree of accuracy and good performance and could be applied to a variety of real-world situations, including surveillance, autonomous vehicles, and augmented reality.

The proposed method can be useful in applications such as robotics [23]. The approach for real-time object detection and classification of small and related figures in image processing was provided in the study. The authors proposed to use image processing techniques to detect and classify objects in images, with a focus on small and similar figures. They evaluated their proposed method on a dataset of images, and reported that the proposed method achieved good performance in the sense of rightness and real-time processing.

## III. PROPSED WORK

### *A. Methodology*

- Dependencies for the model were installed - OpenCV2, Numpy, Mediapipe, Tensorflow.
- Yolov7 model was accessed via the work done in [1] followed by installation of requirement files and necessary packages.
- A python script was built to capture images using a webcam for the dataset.
- A custom dataset consisting of Indian sign language gestures used in day to day lives was created with approximately three thousand images.
- Data annotation was done on the images in the dataset using labelimg.exe - process of labelling data in a format that machines can understand it.
- Defined model configuration and architecture
- Trained custom YOLO v7 detector using 300 epochs.
- Validated the validation set.
- Evaluated custom YOLO v7 performance using TensorBoard.
- Ran inference with trained weights.

### *B. Dataset*

A python script was built to capture images for the dataset. The gestures included in the dataset are - good, very good, good afternoon, school, house and namaste (Indian greeting). Images of about five different people were captured to create a dataset of around thousand images. Further these images were divided into test, validation and train batches which were used to test, validate and train the model respectively.

### *C. Architecture*

The YOLO algorithm stresses only giving an input image one glance. The network is attempting to extract features from the image in order to build a feature map that can be used to recognise items in a particular image while it is being inspected. The network grows more assured in its capacity to foresee items and their classes in a picture as the feature map is filled with more details on the finer characteristics of each object.

A standard convolutional neural network is used for feature extraction. Network's several convolutional layers aid in the transformation of the input image. Bounding box coordinates and object class probabilities are predicted by two fully connected network layers at the pipeline's output.

The architecture can be changed to suit user needs. Typically, this is done by changing the bounding box anchors, the number of training batches, the number of detected objects, and other variables.

## IV. RESULTS

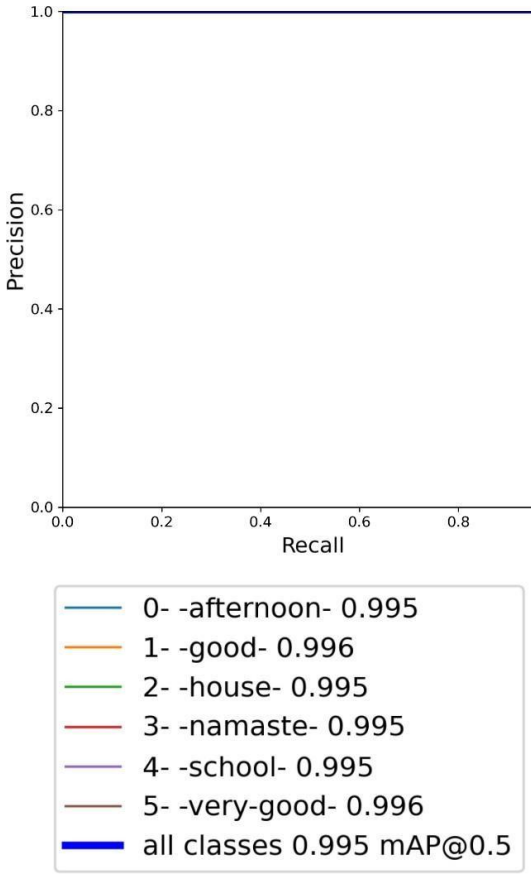


Fig. 3. Precision - Recall curve

Figure 3 shows a precision recall curve. The precision-recall curve shows the trade-off between precision and recall for different thresholds. Low false negative rates are connected with high recall and low false positive rates are correlated with high precision. According to the graph obtained in Figure 3, great recall and high precision are both indicated by a high area under the curve.

After the model was trained, the above results were obtained via Tensorboard. According to the above graph, a perfect precision - recall curve was obtained. It can be safely interpreted that the precision and recall were approximately 99%.

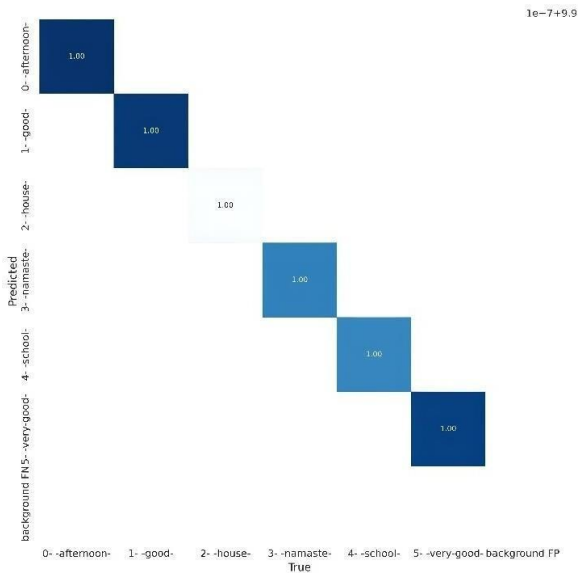


Fig. 4. Confusion Matrix

The efficiency of a classification model is evaluated using a N x N matrix termed a confusion matrix, where N is the total number of target classes. The machine learning model's predicted goal values are compared to the actual goal values in the matrix. In the matrix, the actual goal values are contrasted with those that the machine learning model anticipated. In our study, since six classes were made; the value of N is six.

The confusion matrix further helps calculate the accuracy,

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \qquad \text{Equation 1}$$

Where,
TN = True Negative
TP = True Positive
FN = False Negative
FP = False Positive

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad \text{Equation 2}$$

According to Figure 2, the value of Precision is 1 Therefore, P = 1

According to Equation 2,

$$\frac{\text{TP}}{\text{TP} + \text{FP}} = 1$$

TP = TP + FP

FP = 0

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad \text{Equation 3}$$

According to Figure 2, the value of Recall is 1 Therefore, R = 1

According to Equation 3,

$$\frac{\text{TP}}{\text{FN} + \text{TP}} = 1$$

TP = FN + TP

FN = 0

Equipping Equation 1 with the values of FN (from Equation 2) and FP (from Equation 3),

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN}} = 1$$

Therefore, perfect accuracy was obtained.

Fig. 5. Real Time Recognition of Signs like Namaste and Very Good according to Indian Sign Language

Fig. 6. Real Time Recognition of Signs like Namaste, House, Good and School according to Indian Sign Language

According to the precision - recall graph (Figure 2) and the confusion matrix (Figure 3) it can be concluded that the graph was successfully trained. The model has high precision i.e., 99%.

Figures 5 and Figure 6 illustrate how the model was able to recognise the hand signals in real time with rightness. The model was able to do this easily at the moment even when the lighting was not favorable.

## CONCLUSION AND FUTURE SCOPE

The goal of the study was to develop and put forth a model that could precisely and correctly identify motions in Indian Sign Language. The model was successful in doing so, yolo7 is one of the fastest image detection models and helped gain high precision and accuracy.

Not all words in Indian Sign Language have a single gesture representation, many words are a combination of two or more hand gestures, this model is restricted to detecting words that are only represented by a single gesture.

The model was able to detect two hand gestures as well (example namaste). In future, our study canbe elongated towards words that are a combination of more gestures i.e., rather than detection from an image, detection from a video. This study can help a whole community to communicate easily if the dataset is expanded with more words as the model has high accuracy.

REFERENCES


[1] H. Muthu Mariappan and V. Gomathi, “Real-Time Recognition of Indian Sign Language,” 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1- 6, doi: 10.1109/ICCIDS.2019.8862125.

[2] Rokade, Yogeshwar & Jadav, Prashant. (2017), “Indian Sign Language Recognition System,” International Journal of Engineering and Technology. 9. 189-196. 10.21817/ijet/2017/v9i3/170903S030.

[3] Saurabh Kumbhar, Abhishek Landge, Akash Kulkarni, Devesh Solanki, Vidya Kurtadikar, “Indian Sign Language Recognition System,” June – 2021 International Journal of Innovative Science and Research Technology Volume 6, Issue 6.

[4] Khan A, Sohail A, Zahoora U, Qureshi AS,”A survey of the recent architectures of deep convolutional neural networks,” Artif Intell Rev. 2020;53(8):5455–516.

[5] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[6] Wang, C.-Y., Bochkovskiy, A. & Liao, H.-Y.M,” YOLOv7: Trainable bag-of-freebies sets new state-of- the-art for real-time object detectors,”. Available at: https://arxiv.org/abs/2207.02696v1.

[7] Jacob Solawetz, Joseph Nelson, PP-YOLO Surpasses “YOLOv4 - State of the Art Object Detection Techniques”

[8] Saxena, Rohan and Garg, Romy and Gupta, Bhoomi and Kaur, Narinder, “Hand Gesture Recognition” (February 4, 2022). Available at SSRN.

[9] Raheja, J.L., Mishra, A. & Chaudhary, A. “Indian sign language recognition using SVM.” Pattern Recognit. Image Anal. 26, 434–441 (2016).

[10] Singha, Joyeeta, and Karen Das, "Indian sign language recognition using eigen value weighted Euclidean distance-based classification technique," arXiv preprint arXiv:1303.0634 (2013).

[11] D. Deora and N. Bajaj, "Indian sign language recognition," 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication & Networking, 2012, pp. 1-5.

[12] Bhavadharshini M, Josephine Racheal J, Kamali M, Sankar S, Volume 40: “Advances in Parallel Computing Technologies and Applications”, pp. 159–166.

[13] V. Adithya, R. Rajesh, Hand gestures for emergency situations, “A video dataset based on words from Indian sign language”, Data in Brief, Volume 31, 2020, 106016, ISSN 2352-3409.

[14] Anup Nandy, Jay Shankar Prasad, Soumik Mondal, Pavan Chakraborty, G. C. Nandi, “Recognition of Isolated Indian Sign Language Gesture in Real Time,” Information Processing and Management, 2010, Volume 70.

[15] Bui, Hien Minh, “Hand Sign Language RecognitionWith Artificial Intelligence” Using “You Only Look Once” (Yolo) model as a case 2022 Bachelor Thesis Degree Programme in Business Information Technology Bachelor of Business Administration.

[16] Saxena, R., Garg, R., Gupta, B., Kaur, N. (2023). Deep Learning Based Hand Sign Recognition in the Context of Indian Greetings and Gestures. In: Joby, P.P., Balas, V.E., Palanisamy, R. (eds) IoT Based Control Networks and Intelligent Systems. Lecture Notes in Networks and Systems, vol 528. Springer, Singapore.

[17] Kashish Shah, Sanket Rathi, Rishabh Shetty, Kamal Mistry, “A Comprehensive Review on Text to Indian Sign Language Translation Systems, Smart Trends in Computing and Communications”, 2022, Volume 286.

[18] Ashok Kumar Sahoo, Pradeepta Kumar Sarangi, Chandra Shekhar Yadav, Indian Sign Language Recognition Using Ensemble Based Classifier Combination Macromelecular Synmposia, 10.1002/masy.202100286, 401, 1, (2022).

[19] J. -H. Sun, T. -T. Ji, S. -B. Zhang, J. -K. Yang and G. -R. Ji, "Research on the Hand Gesture

Recognition Based on Deep Learning," 2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE), 2018, pp. 1-4, doi: 10.1109/ISAPE.2018.8634348.

[20] D. Tasmere, B. Ahmed and M. M. Hasan, "Bangla Sign Digits: A Dataset For Real Time Hand Gesture Recognition," 2020 11th International Conference on Electrical and Computer Engineering (ICECE), 2020, pp. 186-189, doi: 10.1109/ICECE51571.2020.9393070.

[21] M. A. Rahim, J. Shin and M. R. Islam, "Dynamic Hand Gesture Based Sign Word Recognition Using Convolutional Neural Network with Feature Fusion," 2019 IEEE 2nd International Conference on Knowledge Innovation and Invention (ICKII), 2019, pp. 221-224, doi: 10.1109/ICKII46306.2019.9042600.

[22] C. Liu, Y. Tao, J. Liang, K. Li and Y. Chen, "Object Detection Based on YOLO Network," 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), 2018, pp. 799-803, doi: 10.1109/ITOEC.2018.8740604.

[23] A. M. Algorry, A. G. García and A. G. Wofmann, "Real- Time Object Detection and Classification of Small and Similar Figures in Image Processing," 2017 International Conference on Computational Science and Computational Intelligence (CSCI), 2017, pp. 516-519, doi:10.1109/CSCI.2017.8.