Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 05, May-2025

Recognising Words in American Sign Language: A YOLOv11 Based Approach

Milan Singhal
School of Computer Science
University of Petroleum and
Energy Studies
Dehradun, India

***Abstract*— By using a live camera feed, this project tries to translate ASL into understandable sentences. A main part of the research is to develop a live translation tool to connect people relying on sign language. YOLO is used in our approach so that video input is analyzed in real time, without needing any prior learning targeted to a specific area. Our study is one of the few that apply YOLO and we believe it demonstrates how this architecture can support effective and real-time translation of sign language.**

***Keywords*— Sign Language Recognition; You Only Look Once (YOLO); American Sign Language, Natural Language Processing; Machine Learning; Deep Learning, Computer Vision**

I. INTRODUCTION

People use language to interact, exchange thoughts, feelings and intention with each other. Spoken language usually creates a big obstacle for those who are deaf or hard-of-hearing. For this reason, sign language has grown as the main way to communicate, using hand movements, body shapes and facial expressions. Although there have been big improvements in communication methods, many people who use sign language still face gaps with those who use spoken or written language. To connect this gap, smart systems must be developed that interpret sign language as it happens, promoting no-barrier and inclusive talk.

Hand gestures, signals, body movements, facial expressions, and lip movements are the visual means of communication used by the deaf and hard-of-hearing community. This language is identified as sign language. Sign language recognition (SLR) is a very challenging and complex task for better communication, it can be done using various research opportunities available with artificial intelligence and deep learning. SLR aims to recognize and understand sign gestures by suitable and efficient techniques, which requires identifying the features and classifying the sign as gesture recognition.

According to WHO (World Health Organization), over 5% of the world's population, or 430 million people which includes 34 million children, require rehabilitation for their disabling hearing loss treatment. It is estimated that by 2050 over 700 million individuals or 1 in every 10 people will have disabling hearing loss. Nearly 80% of people with disabling hearing loss originate from low and middle-income countries. The frequency of hearing loss increases with age, among those older than 60 years, it has increased by 25%.

In India, there are significant divergences in the number of hearing-impaired individuals. According to the 2011 Census report, there are five million deaf and hard-of-hearing communities, 18 million according to the National Association of the Deaf, and nearly 63 million according to the World Health Organization. Despite these numbers, only 5% of deaf children are in school, and deaf adults struggle to secure employment.

II. RELATED WORKS

Here are some of the selected research works that inspired us to start working on our thesis topic in depth. First, we will discuss the works related to text to sign language translation. In a unique approach to translating English sentences to Indian Sign Language (ISL) is seen. Their proposed system takes a text input and converts it to ISL with the help of Lexical Functional Grammar (LFG). In an approach to transform Malayalam text to Indian Sign Language using animation for displaying is seen. Their system uses the Hamburg Notation System shortly known as HamNoSys for representing signs. Moreover, the authors in used an approach for converting Greek text to Greek sign language. Translation is done using V signs, a web tool used for the synthesis of virtual signs. A system is proposed where text in English language is taken as input and then translated to HamNoSys representation. This is afterward converted into SiGML. A mapping system is used to link the text to the HamNoSys notation. This work may not be a direct example of text to-sign language conversion which we expect. However, this provides us with insights into converting text to a signed notation system. Similar research works were done in and furthermore, in the authors proposed a machine translation model that takes both examples based and rule-based Interlingua approaches to convert Arabic Text to Arabic Sign Language. Another work of Arabic Sign language for the deaf is presented. In Adding to that, in a text-to-sign language conversion system for Indian Sign Language (ISL) is made which takes into account the language's distinctive alphabet and syntax. The system accepts input in alphabets or numerals only.

Now, we will discuss the works related to sign language recognition. In the authors attempted to recognize the English alphabet and gestures in sign language and produced the accurate text version of the sign language using CNN and computer vision. In the researchers worked on reviewing multiple works on the recognition of Indian Sign Language

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 05, May-2025

(ISL). Their review of works on Histogram of Orientation Gradient (HOG), Histogram of Edge Frequency (HOEF) and Support Vector Machine (SVM) gave us meaningful insights. A similar work is seen in Furthermore, in the authors worked on Indonesian sign language recognition was done using a YOLOv3 pre-trained model. They used both image and video data. The system's performance was incredibly high during using image data and it was comparatively low while using video data. A similar work was done in using YOLOv3 model. From we learnt how the researchers worked on making an Italian sign language recognition system that identifies letters of the Italian alphabet in real-time using CNN and VGG-19. The work of the authors in and was insightful about how deep learning works on sign language detection. Moreover, the authors developed an Android app that can convert real-time ASL input to speech or text where SVM was used to train the proposed model. Additionally, in we were introduced to the idea of using surface electromyography (sEMG), accelerometer (ACC), and gyroscope (GYRO) sensors for sub word recognition in Chinese Sign Language. Lastly in the authors worked on a sign language-to-voice turning system that uses image processing and machine learning.

III. METHODOLOGY

A. Sign Language To Text

Using YOLOv8 and YOLO11, we have developed two models for detecting sign language. YOLOv11 is the newest and best YOLO model that stands for the latest discoveries in the series. The new YOLO model provides additional advantages in detecting objects over its previous versions. The model trains using images and can precisely recognize gestures and other types of situations. The custom data loader mosaic in YOLO11 streamlines and accelerates both learning and testing using raw data.

Even so, the way data is gathered is much the same for YOLOv8 as it is for YOLO11, since they both have similar structures. Many data collection strategies work with both models. We chose bounding box prediction, as our movements involve a lot of varying shapes and are not always straight. Therefore, we set our hand gestures inside bounding boxes, because this makes the data easier to handle and the results been reasonably good for both models.

Dataset preparation: We have used a word corpus that was custom designed for ASL research. There are 18 different classes in the dataset out of which there are various words, including call, dislike, fist, four, like, mute, okay, one, palm, peace, peace inverted, rock, stop, stop inverted, three, three 2, two up and two up inverted. All the images were made 640×480 in size and labeled using the LabelImg tool following their class.

In this project, the dataset is separated into Training, Validation and Testing sets, making up 64%, 16% and 20%, respectively.

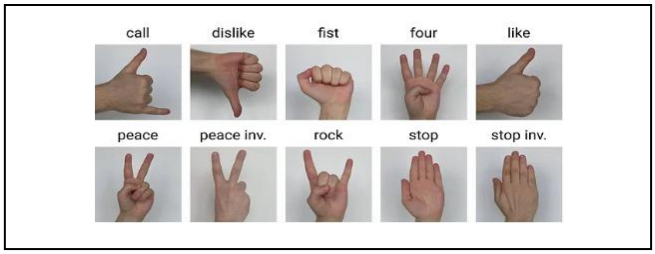

fig. 1 – Images of some training samples

Furthermore, most often the data of a single word spells have some movement including hand gestures. Therefore, we used samples which include movement for those specific characters. The accuracy remained high as we used movement gestures. We took 350 image samples for each class in the training data, 25 image samples for each class in testing data and approximately 30 images in validation data. Overall, we got more than 10000 data samples, where we used 9450 samples for training, 775 samples for testing our data and 700 images for validating the dataset.

Implementation of YOLO models:

In the beginning, we went over each sample one after another, using custom code to give each a class and a number. Afterwards, we assigned each of the sample images to their correct classes so real-time annotation could begin. We used the same data we collected and set up the model to run through our chosen epochs. Our model was trained on 64 percent of the created information, tested on 20 percent and validated with what remained, at 16 percent. After training both times, we took the model that fit best and got the same result.

Test results of YOLOv11 Model:

Here, 50 epochs were executed on YOLO, with a total of 18 classes. In fig. 2 below, real time detection of some classes is shown. From fig. 3, we can see Precision-Recall Curve of the model discussed.

fig. 2 – Images of some training samples

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 05, May-2025

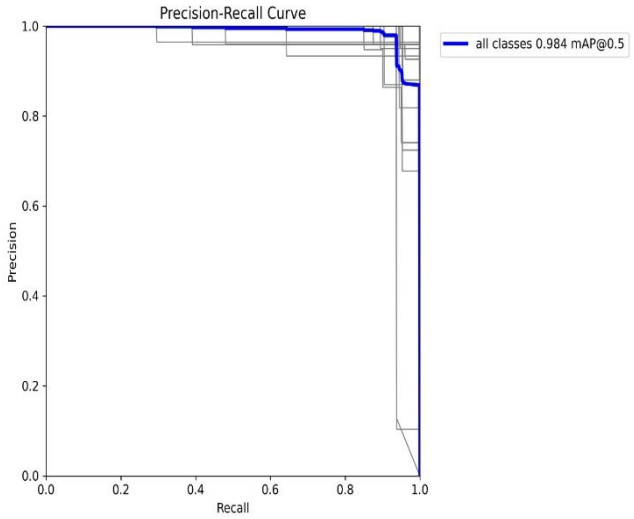

fig. 3 – Precision Recall Curve

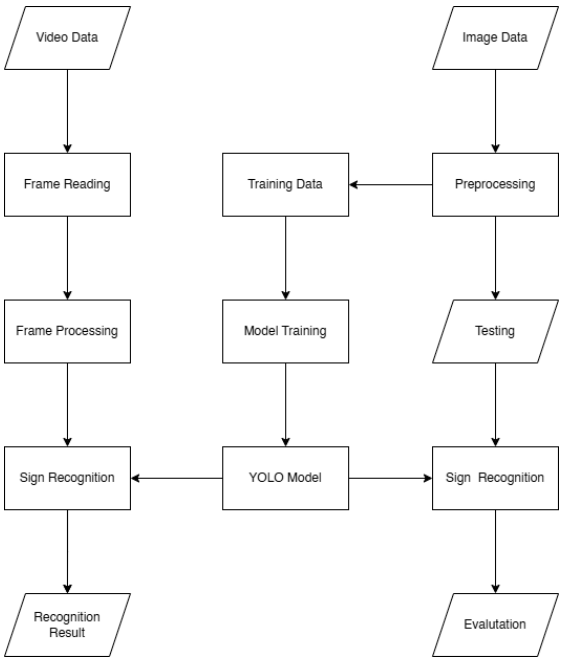

fig. 4 – Process Flow

IV. PROCESS FLOW

Figure 4 demonstrates the complete process that our sign language recognition system follows. Input is provided to the process by two sources: a video and an image. For video clips, we take out single frames and treat them to distinguish the necessary gesture. All images are standardized and formatted when image data is processed. From here, both types of data are separated into training and testing sets. YOLO is trained with the training data to recognize and classify movement for 18 distinct gesture categories. After training, the model is tested on data that hasn't been seen before. Following recognition, the outcomes are assessed to see if the model is accurate and reliable. At this point, fundamental evaluation metrics such as the confusion matrix, are produced. Also, the verified model is put to use for live gesture recognition. This structured method correctly detects and classifies gestures with the help of state-of-the-art deep learning methods.

V. EVALUATION OF RESULT

From the table in fig. 5 it is visible that the YOLO v8 with 16 batch size shows good accuracy, but sometimes it makes false detection. The model is unable to always perfectly identify the hand signs where both hands have been used to perform the sign despite providing enough data. So, it is difficult for this model to identify or recognize comparatively complex hand sign gestures. After seeing its performance, we can say that this model is not suitable for deployment purposes at large scale work places.

The YOLO v11 model with the same batch size on the other hand have performed better than the other model. Though the this model shows a little bit less accuracy but it does not make any type of false detection in presence of multiple hands togather in a single frame, instead it percieves those signs more accurately and tries to predict the actual words more accurately.

Model	Classes	Batch Size	Epoch	Accuracy
YOLO v8	18	16	50	97.2
YOLO v11	18	16	50	96.6

fig. 5 – Performance Comparison

VI. CONCLUSION

In the text to sign language conversion framework, there are certain sentences that contain stop words (For example- 'apostrophe) that we utilized for filtering are not compatible with the framework. In future we can also incorporate a 3D model with smoother transitions. Moreover, training a model over the video dataset, thus increasing it will also let us reach new horizons of the research and later on adding some facial action reaction recognitions for better understanding of the semantics. In near future we can also plan to make an app

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 05, May-2025

version of this model and framework too. On top of it, we state that our work here on ASL detection can also be applied to other sign languages as well. According to the World Health Organization (WHO), with 1.5 billion people in the world already suffering from hearing loss and the number can increase to over 2.5 billion by 2050. The deaf community is deprived of basic human rights like health care, education and even minimum wage jobs simply because of their inability to communicate with the hearing people using spoken language. This YOLO based model and the NLP based framework aim to bridge this communication gap that is prevalent in the community for a long time by providing the fastest real time solution. This will ensure an equal spot for the deaf people in the society by overcoming the language barrier. In conclusion, this system will be helpful for both hearing- and hearing-impaired people to communicate effectively with one another by shortening the existing communication gap.

VII. REFERENCES

[1] T. Dasgupta, S. Dandpat, and A. Basu, "Prototype machine translation system from text-to-Indian sign," NLP for Less Privileged Languages, vol. 19, 2008.

[2] Yoav Goldberg, "Sign Language Processing — research.sign.mt," [On line]. Available: https://research.sign.mt. [Accessed 25-May-2023].

[3] J. Singh and D. Singh, "Sign language and hand gesture recognition using machine learning techniques: A comprehensive review," Modern Computational Techniques for Engineering Applications, pp. 187–211, CRC Press.

[4] S. Daniels, N. Suciati, and C. Fathichah, "Indonesian sign language recognition using YOLO method," in IOP Conference Series: Materials Science and Engineering, 2021, vol. 1077, no. 1, p. 012029.

[5] MAMM Asri, Zaaba Ahmad, Itaza Afiani Mohtar, and Shafaf Ibrahim, "A real-time Malaysian sign language detection algorithm based on YOLOv3," International Journal of Recent Technology and Engineering, vol. 8, no. 2, 2019, pp. 651–656.

[6] M. S. Nair, A. P. Nimitha, and S. M. Idicula, "Conversion of Malayalam text to Indian sign language using synthetic animation," in 2016 Inter national Conference on Next Generation Intelligent Systems (ICNGIS), 2016, pp. 1–4.

[7] D. Kouremenos, S.-E. Fotinea, E. Efthimiou, and K. Ntalianis, "A prototype Greek text to Greek Sign Language conversion system," Behaviour & Information Technology, vol. 29, no. 5, 2010, pp. 467 481.

[8] M. Varghese and S. K. Nambiar, "English to SiGML conversion for sign language generation," in 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), 2018, pp. 1–6.

[9] K. Kaur and P. Kumar, "HamNoSys to SiGML conversion system for sign language automation," Procedia Computer Science, vol. 89, 2016, pp. 794–803.

[10] V. J. Schmalz, "Real-time Italian Sign Language Recognition with Deep Learning," in CEUR Workshop Proceedings, 2022, vol. 3078, pp. 45–57.

[11] Z. Alsaadi, E. Alshamani, M. Alrehaili, A. Alrashdi, S. Albelwi, and A. O. Elfaki, "A real-time Arabic sign language alphabets (ArSLA) recognition model using deep learning architecture," Computers, vol. 11, no. 5, 2022, p. 78.

[12] X. Yang, X. Chen, X. Cao, S. Wei, and X. Zhang, "Chinese sign language recognition based on an optimized tree-structure framework," IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 4, 2016, pp. 994–1004, IEEE.

[13] S. Sudeep, "Text to Sign Language Conversion by Using Python and Database of Images and Videos," www.academia.edu. [Online]. Available: https://www.academia.edu/40531482/TexttoS.

[14] A. Kamble, J. Musale, and R. Chalavade, "Conversion of Sign Language to Text," www.ijraset.com, May 2023. [Online]. Available: https://www.ijraset.com/research-paper/ conversion-of-sign-language-to-text#introduction.

[15] Prof. C. Narvekar, S. Mungekar, A. Pandey, and M. Mahadik, "SIGN LANGUAGE TO SPEECH CONVERSION USING IMAGE PRO CESSING AND MACHINE LEARNING," Aug. 2020. [Online]. Available:https://ijtre.com/wp-content/uploads/2021/10/2020071224.pdf.

[16] A. Singh Dhanjal and W. Singh, "An automatic conversion of Punjabi text to Indian sign language," EAI Endorsed Transactions on Scalable Information Systems, vol. 7, no. 28, 2020, pp. e9–e9.

[17] M. Brour and A. Benabbou, "ATLASLang MTS 1: Arabic text language into Arabic sign language machine translation system," Procedia Com puter Science, vol. 148, 2019, pp. 236–245, Elsevier.

[18] K. Tiku, J. Maloo, A. Ramesh, and R. Indra, "Real-time conversion of sign language to text and speech," in 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 346–351, IEEE.

[19] S. Dangsaart, K. Naruedomkul, N. Cercone, and B. Sirinaovakul, "Intel ligent Thai text–Thai sign translation for language learning," Computers & Education, vol. 51, no. 3, 2008, pp. 1125–1141, Elsevier.

[20] R. Alzohairi, R. Alghonaim, W. Alshehri, and S. Aloqeely, "Image based Arabic sign language recognition system," International Journal of Advanced Computer Science and Applications, vol. 9, no. 3, 2018, Science and Information (SAI) Organization Limited.

[21] World Health Organization, "Hearing Loss." [Online]. Available: https: //www.who.int/health-topics/hearing-loss#tab

[22] Lang S, Block M and Rojas R 2012 Sign language recognition using kinect Int. Conf. on Artificial Intelligence and Soft Computing pp 394–402