

Sign Language Recognition

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

Abstract— The primary concern of this project is to take American Sign Language (ASL) data through live camera feed and convert those letters (alphabets) and information into plain text. Additionally, this research also focuses on creating a framework that could help converting the sign language in real time thus breaking the language barrier for the people in need. For the research, we have used You Only look Once (YOLO) model, executing in real time extracting discriminating spatial-temporal data without any prior knowledge in the domain. To the best of our knowledge, this is a rare study using only YOLO model to demonstrate bidirectional sign language communication in real time in the American Sign Language (ASL).

Keywords— Sign Language; You Only Look Once; American Sign Language, Natural Language Processing; Machine Learning; Deep Learning

million individuals or 1 in every 10 people will have disabling hearing loss. As you can see in fig 1, which tells us about disabling hearing loss that refers to hearing loss greater than 35 decibels (dB) in the better-hearing ear. 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.

I. INTRODUCTION

Language serves as the basic foundation for human communication and interaction. It enables everyone to express thoughts, emotions, and intentions. However, for the deaf and hard-of-hearing community, traditional verbal communication methods become a significant barrier, so this community use sign language for their interaction and communication. Even though communication has advanced significantly in this era, there remains a crucial gap between sign language users and those who rely on spoken or written language for smooth communication. Bridging this gap requires innovative technological solutions that can accurately interpret sign language accurately and efficiently.

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

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 (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

1) YOLO Models: For our sign language detection, we have developed models based on YOLOv5 and v8. YOLOv8 is a state-of-the-art model in the ongoing time. This YOLO model is an upgraded version of its predecessors. It uses image data to train and can detect the gestures and other detection scenarios with exceptional accuracy. YOLOv8 uses a custom data loader mosaic which helps the data to be loaded into the model while training and testing the raw data. However, the data collection is more likely the same with v5 and v8 models as both have a similar architectural pattern. Both of these models are compatible with multiple data collection styles. We used bounding box prediction data style as it was more suitable for us as our hands move constantly and we will have to deal with more than one shape and aspect ratios. Therefore, we added our hand gestures with bounding boxes as it can be easily manipulated and the end result was satisfactory for both of the models.

Dataset preparation: We have used ASL Dataset. In this dataset, there are 27 classes in total, (A-Z and '-' (nothing)). After gathering images for each class we resized them at 640*480px and annotated each of the data samples according to their specific class using 'LabelImg' tool.

Further the classes are divided as 80 (Training) : 10 (Testing) : 10 (Validation). Below in fig. 1 are some sample images of the classes used



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, 675 samples for testing our data and 800 images for validating the dataset.

Implementation of YOLO models:

At first, we took the sample data one by one using a custom code after assigning the classes and numbers. Afterwards, we annotated each of the sample pictures with its corresponding class so that the actual annotation in real time may work. We preprocessed the same data which we collected and ran our decided epochs for the model. Our data was trained with 80 percent of the created dataset, tested with 10 percent of it and validated with the rest 10 percent. We took the best fit model we got from both of the training sessions and got the desired output.

Test results of YOLOv8 Model:

Here, 50 epochs were executed on YOLO, with a total of 27 classes. In fig. 2 below, real time detection of some classes is shown. From fig. 3, we can see the F1 Score and in fig. 4 Confusion Matrix created for the above discussed model.

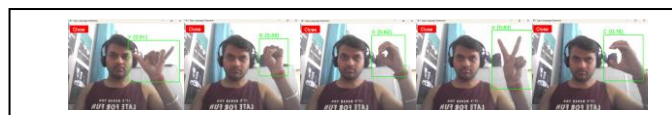


fig. 2 – Images of some training samples

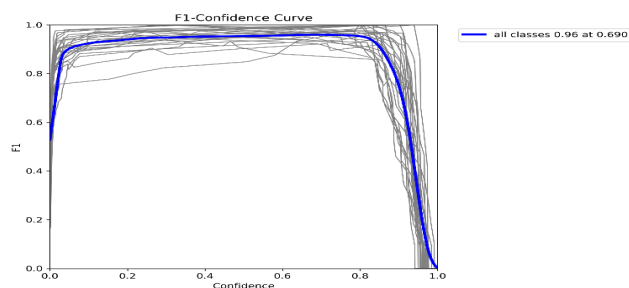


fig. 3 – F1 Score

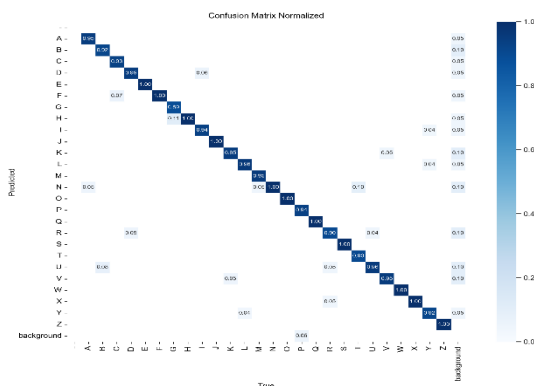


fig. 4 – Confusion Matrix

IV. EVALUATION OF RESULT

From the table in fig. 5 it is visible that the YOLO v8 with 1 batch size shows a high 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 have come to the conclusion that this model is not suitable for large scale work.

The YOLO model with batch size of 16 on the other hand have performed much better than the other model. Though the this model shows a little bit less accuracy but it does not do any type of false detection. The YOLOv8 model also does not have any case of making any wrong detection.

Model	Classes	Batch Size	Epoch	Accuracy	F1 Score
YOLO v8	27	1	50	97.2	0.71
YOLO v8	27	16	50	95.6	0.65

fig. 5 – Performance Comparison

V. 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 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.

VI. 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," [Online]. 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 International 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 (ICSDDET), 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. Alshamami, 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 PROCESSING 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 Computer 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. Dangsart, 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>