Journal of Applied Science, Engineering, Technology and Management
Vol.01, Issue 02, December 2023 pp. 13 - 16

Sign Language Recognition and Video Generation Using Deep Learning

Meera Treesa Mathews
Department of Computer Science and Engineering
Federal Institute of Science and Technology
Angamaly, Kerala, India
meeratreesa@fisat.ac.in

Joyal Raphel
Department of Computer Science and Engineering
Federal Institute of Science and Technology
Angamaly, Kerala, India
joyalraphel100@gmail.com

Joseph Shaju C
Department of Computer Science and Engineering
Federal Institute of Science and Technology
Angamaly, Kerala, India
josephshaju123@gmail.com

Steve Soney Varghese
Department of Computer Science and Engineering
Federal Institute of Science and Technology
Angamaly, Kerala, India
stevesoneyvarghese13@gmail.com

Paul J Puthussery
Department of Computer Science and Engineering
Federal Institute of Science and Technology
Angamaly, Kerala, India
pauljp04@gmail.com

***Abstract* — The proposed system aims to help normal people understand the communication of speech impaired individuals through hand gestures recognition and generating animation gestures. The system focuses on recognizing different hand gestures and converting them into information that is understandable by normal people. YOLOv8 model, a state-of-the-art object detection algorithm, is being employed in this system to detect and classify sign language gestures. Sign language video generation can act as a guide for anyone who is in the process of learning sign language, by providing them with expressive sign language videos using avatars that can translate the user inputs to sign language videos. CWASA Package and SiGML files are used for this process. The project contributes to the advancement of assistive technologies for the hearing-impaired community, offering innovative solutions for sign language recognition and video generation.**

Keywords — Gesture recognition, Deep Learning, Sign language, Video generation.

I. Introduction

Sign language is the bridge that connects us to the deaf world. It also helps deaf people communicate with people around them. You can understand the world around you through visual descriptions and contribute to society. The language relies primarily on hand articulations and non-manual gestures [1]. Ordinary people have a hard time understanding their own language. Therefore, we need a system that recognizes various signs and gestures and conveys information to the general public. Sign language recognition systems help recognize signs and translate them into corresponding words. It bridges the gap between disabled and normal people [2-5]. People who are learning sign language for the first time may feel overwhelmed by the learning process. It might be difficult to have to remember the signs corresponding to each word. In such situations it might be helpful for them to have a system that could help them in the learning process by acting as a guide, by showing them the sign representations for each word that they are having trouble with, in an interactive way. Our system puts forward an end-to-end deep learning framework to efficiently manage processing of the sign language. The Sign Language recognition model is developed using the YOLO model. The dataset is labelled and annotated using Roboflow and trained using YOLO model. Certain pre-processing and post- processing techniques are adopted to guarantee high efficiency. The SL recognition provides an accuracy of 90% in this system. For video generation, the user input text is first converted to ISL form using NLP techniques and tools such as Stanza, Stanford Parser etc. It is then converted to avatar animations using CWASA Packageand SiGML files [6-8]. The project contributes to the advancement of assistive technologies for the hearing-impaired community, offering innovative solutions for sign language recognition andvideo generation.

II. Problem Statement

Communication is the act of giving, receiving, and sharing information, but this is very difficult for specially disabled like deaf and mute. Specially-abled people use hand signals to communicate, which makes it difficult for ordinary people to recognize their own words from sign language [9,10]. We need a system to tell for people who are learning sign language they feel it difficult to represent normal sentence in signs. Here sign language video generation help them in teaching how to properly generate sign from normal sentences [11,12].

eISSN: 2584-0371

Received for review: 16-08-2023
Accepted for publication: 21-11-2023
Published: 02-12-2023

III. METHODOLOGY

The project is designed to identify words for corresponding sign gestures and to produce sign gesture animation for corresponding input sentences. So, it got two parts Sign Language Recognition and Video Generation.

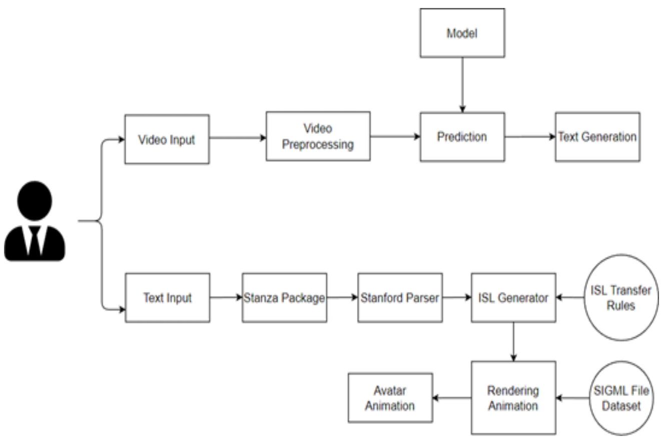

Fig. 1 System Architecture

A. *Sign Language Recognition*

1) Data Collection: The design methodology begins with a comprehensive data collection process, we have collected datasets for different Indian sign language [13] representation like letters for English alphabets, numbers and for some common expressions like "Thank You", "Please", "Yes" etc [14]. We have created at most 1500 dataset for training.

2) Dataset Labeling: The generated images was imported to Roboflow and the datasets was labeled and classified each image to corresponding classes. The datasets were subsequently split into testing, training, and validation datasets. For scaling purpose, we used min max scaler from the SKlearn.

3) Model Training: The model was trained in Google Colab. We employed the YOLO v8 algorithm to train our sign language recognition model. We have tried various YOLO v8 models like YOLOv8n, YOLOv8m, YOLOv8l etc. and decided to stick on small model as it will good for prediction of sign language at real time.

4) Model Optimization: The training process involved several iterations, gradually refining the model's performance. We trained the model at various epochs, evaluating the mean average precision and loss function to assess its progress. Ultimately, we concluded the training after 200 epochs, as the model achieved satisfactory results.

5) Prediction and Output: The generated weights were downloaded for sign language recognition. The webcam continuously captured images, which were then processed by the system to identify sign gestures. Upon detection, the model predicted the corresponding sign and plotted a bounding box around it. The label of the recognized sign was displayed above the bounding box, along with the corresponding confidence score, indicating the model's certainty in its prediction.

B. *Sign Language Video Generation*

1) Model Selection: Since stanza provides models for variety of languages, it was important to select basic language model. We chose the English NLP (Natural Language Processing) model and passed the input to it. The model returns a property detailing each word and corresponding sentences briefly.

2) ISL Conversion using Parse Tree: We represented each sentence in the form of annotated parse tree (provided by Stanford Parser). So now we modify and reorder the parse tree with the help of ISL grammar rules. The word list gets converted to ISL text format.

3) Preprocessing: The parsed ISL text need to be further altered by removing stop words, replacing words with corresponding lemma, removing punctuation. After the completion of pre-processing the final list need to be verified with the SiGML files available.

4) SiGML File Collection: We downloaded various SiGML files available for corresponding words. In case if a word is not available, it is replaced with each of its letters in the final list. We collected over 850 SiGML files including alphabets and numbers.

5) Animation Output: The final list is then rendered with the help of CWASA. CWA Signing Avatars provides the avatar definition and acts like a SiGML player that plays the SiGML file word by word hence producing the final animation output.

IV. RESULTS AND DISCUSSION

A. *Sign Language Recognition*

Fig. 2 Sign language recognition output

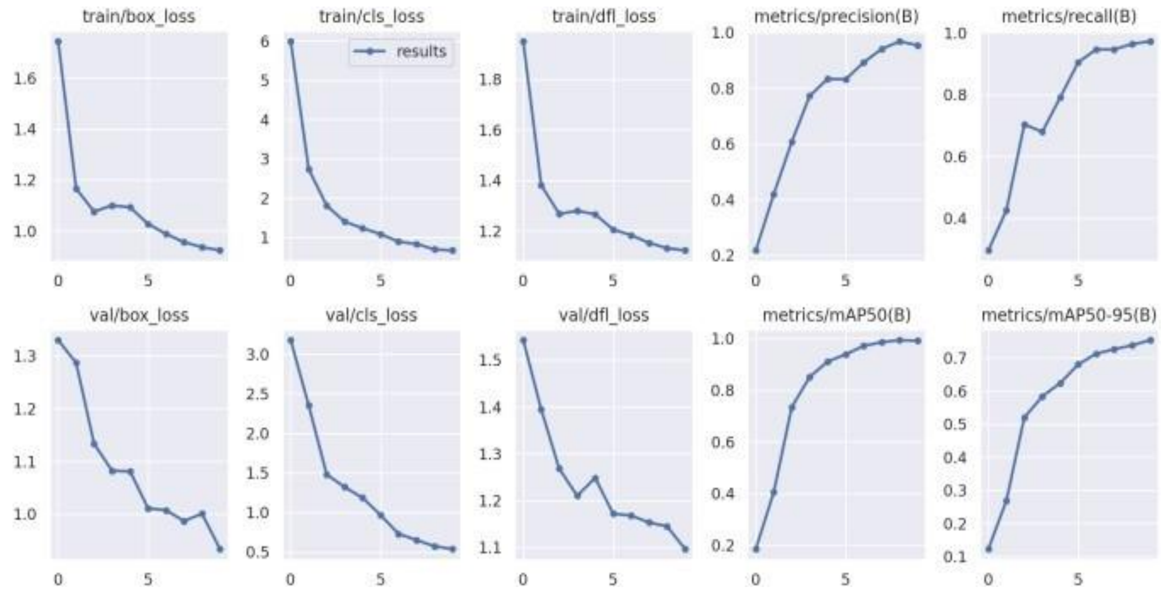

Fig. 3 Performance Metrics

As shown in Figure 2, Sign Language Recognition system continuously captures images, which were then processed by the system to identify sign gestures. Upon detection, the model predicted the corresponding sign and plotteda bounding box around it. The label of the recognized sign was displayed above the bounding box, along with the corresponding confidence score, indicating the model's certainty in its prediction.

Figure 3 show various values associated with model like loss function, mean average precision matrix, recall etc... These values were taken into consideration while selecting the number of iterations required for training the model.

B. Sign Language Video Generation

Figure 4 represents the video generation output. As the figure shows, the ISL format text is generated once the user enters his input and presses the "Submit" button. The animation rendering takes place when the "Render Animation" button is pressed. This section includes the evaluation of the model trained throughout project and other results.

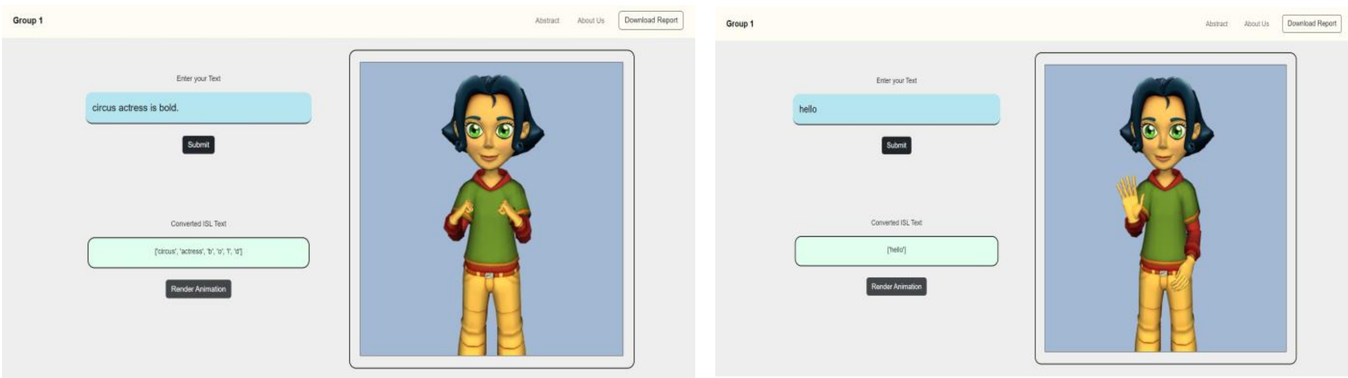

Fig. 4. Video generation output

V. Conclusion

This work was able to successfully implement a comprehensive framework for sign language recognition and video generation using deep learning techniques. By leveraging the YOLOv8 model for sign language recognition, using NLP techniques for general English to sign language text conversion and integrating the CWASA Package and SiGML files for video generation, we have addressed the communication challenges faced by individuals with hearing impairments. Through extensive training on annotated datasets and fine-tuning of the YOLOv8 model, sign language recognition system achieves accurate and efficient recognition of sign language gestures. This breakthrough empowers individuals with hearing impairments to communicate effectively with the broader community, fostering inclusivity and bridging the communication gap. Furthermore, the avatar animation system generates expressive sign language videos based on user input, facilitating the learning process and providing a valuable guide for individuals new to sign language. By visualizing sign language representations in an interactive and dynamic manner, our system enhances accessibility and promotes a deeper understanding of sign language for both learners and the wider community. The relevance of our project lies in its potential to transform the way we perceive and engage with sign language. By combining deep learning, computer vision, and animation techniques, we have made significant strides in improving communication, accessibility, and inclusivity for individuals with hearing impairments. While the project has achieved notable success, there are still opportunities for future enhancements. Exploring advanced architectures, incorporating multi-modal integration, and enabling real-time translation are among the potential areas for further research and development [15]. This project represents a significant contribution to the field of sign language recognition and video generation. It underscores the importance of leveraging deep learning technologies to break down communication barriers and foster a more inclusive society.

References

[1] H. Luqman, "An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion," in IEEE Access, vol. 10, pp. 93785-93798, 2022, doi: 10.1109/ACCESS.2022.3204110.

[2] Gamal Tharwat, Abdelmoty M. Ahmed and Belgacem Bouallegue, "Arabic Sign Language Recognition System for Alphabets Using Machine Learning Techniques;(IEEE Access 2019)

[3] Walaa Aly and Saleh Aly, "DeepArSLR: A Novel Signer-Independent Deep Learning Framework for Isolated Arabic Sign Language Gestures Recognition'; (IEEE Access 2020)

[4] F. Shah, M. S. Shah, W. Akram, A. Manzoor, R. O. Mahmoud and D. S. Abdelminaam, "Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language," in IEEE Access, vol. 9, pp. 67548-67558, 2021, doi: 10.1109/ACCESS.2021.3077386.

[5] S. K. Youme, T. A. Chowdhury, H. Ahamed, M. S. Abid, L. Chowdhury and N. Mohammed, "Generalization of Bangla Sign Language Recognition Using Angular Loss Functions," in IEEE Access, vol. 9, pp. 165351- 165365, 2021, doi: 10.1109/ACCESS.2021.3134903.

[6] M. D. Nguyen, H. N. Vu, D. C. Pham, B. Choi and S. Ro, "Multistage Real-Time Fire Detection Using Convolutional Neural Networks and Long Short-Term Memory Networks," in IEEE Access, vol. 9, pp. 146667-146679, 2021, doi: 10.1109/ACCESS.2021.3122346.

[7] B. Zhang and Y. Bao, "Cross-Dataset Learning for Age Estimation," in IEEE Access, vol. 10, pp. 24048-24055, 2022, doi: 10.1109/AC-CESS.2022.3154403.

[8] J. R. Castillo and M. J. Flores, "Web-Based Music Genre Classification for Timeline Song Visualization and Analysis," in IEEE Access, vol. 9, pp. 18801-18816, 2021, doi: 10.1109/ACCESS.2021.3053864.

[9] M. Al-Hammadi et al., "Deep Learning-Based Approach for Sign Language Gesture Recognition with Efficient Hand Gesture Representation," in IEEE Access, vol. 8, pp. 192527-192542, 2020, doi: 10.1109/ACCESS.2020.3032140.

[10] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif and M. A. Mekhtiche, "Hand Gesture Recognition for Sign Language Using 3DCNN," in IEEE Access, vol. 8, pp. 79491-79509, 2020, doi: 10.1109/ACCESS.2020.2990434.

[11] B. Natarajan et al., "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation," in IEEE Access, vol. 10, pp. 104358-104374, 2022, doi: 10.1109/ACCESS.2022.3210543.

[12] Y. Liao, P. Xiong, W. Min, W. Min and J. Lu, "Dynamic Sign Language Recognition Based on Video Sequence With BLSTM-3D Residual Networks," in IEEE Access, vol. 7, pp. 38044-38054, 2019, doi: 10.1109/ACCESS.2019.2904749.

[13] J. Joy, K. Balakrishnan and M. Sreeraj, "SignQuiz: A Quiz Based Tool for Learning Fingerspelled Signs in Indian Sign Language Using ASLR," in IEEE Access, vol. 7, pp. 28363-28371, 2019, doi: 10.1109/AC- CESS.2019.2901863.

[14] W. Aly, S. Aly and S. Almotairi, "User-Independent American Sign Language Alphabet Recognition Based on Depth Image and PCANet Features," in IEEE Access, vol. 7, pp. 123138-123150, 2019, doi: 10.1109/ACCESS.2019.2938829.

[15] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate" in arXiv:1409.0473 [cs.CL].