2024 International Conference on Communication, Computing and Internet of Things (IC3IoT) | 979-8-3503-5268-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/IC3IoT60841.2024.10550211

Animated Sign Language For People With Speaking And Hearing Disability Using Deep Learning

R Valarmathi
Department of Computer Science and Engineering
Sri Sairam Engineering College
Chennai, India.
valarmathi.cse@sairam.edu.in

Jaya Surya P
Department of Computer Science and Engineering
Sri Sairam Engineering College
Chennai, India.
sec21cs081@sairamtap.edu.in

Balaji P
Department of Computer Science and Engineering
Sri Sairam Engineering College
Chennai, India.
sec21cs167@sairamtap.edu.in

Ashik K
Department of Computer Science and Engineering
Sri Sairam Engineering College
Chennai, India.
sec21cs090@sairamtap.edu.in

***Abstract*— This paper addresses the communication barriers faced by individuals with speaking and hearing disabilities through the utilization of deep learning techniques. Our System consists of two modules where one converts the sign gestures performed by the deaf and dumb people to audio and the second module converts the words spoken by the normal people into Animations which perform the sign gestures corresponding to the audio. To enhance the user experience, we have integrated natural language processing (NLP) techniques to improve the fluency and coherence of the translated text. Additionally, our system utilizes expressive 3D gestures to animate the sign gestures, making the communication more engaging and relatable. These gestures are customizable to match the user's identity, further personalizing the interaction. Our system leverages the power of deep learning, NLP, and 3D animation to break down communication barriers for individuals with hearing and speaking disabilities. This innovative solution has the potential to transform the lives of millions by providing a means for them to effectively communicate with the broader society, thereby fostering inclusivity and equal opportunities.**

Keywords—Communication Barriers, Natural Language Processing (NLP), Sign Gestures, Techniques in Deep Learning, User Experience.

I. Introduction

In contemporary society, accessibility and inclusivity have gained paramount importance, and technology has emerged as a potent tool in bridging the problems faced by the individuals with listening and speaking disabilities. The disabled community, constituting millions worldwide, often face significant challenges in their daily lives, primarily related to the communication barrier between themselves and those who rely on spoken language. Traditional sign language interpretation typically relies on human interpreters, creating challenges such as limited availability, varying interpretation skills, and cost constraints.

However, recent advancements in deep learning and computer vision technologies offer a promising avenue to address these issues by automating the process of sign language interpretation and rendering it in the form of animated gestures. This paper introduces a novel and innovative approach to empower individuals with hearing and speech impaired disabilities through the deep learning techniques to create an animated sign language Mechanism. Our research primarily aims to develop an automated system that can recognize and interpret sign gestures with high accuracy and convert them audio and similarly the words spoken by the normal to into animated gestures. These gestures can serve as intermediaries, allowing deaf individuals to communicate more seamlessly with the hearing community and vice versa. Additionally, this technology can assist individuals with speaking disabilities by providing them with a means to express themselves more fluently. This paper presents a complete exploration of sign language recognition, addressing the unique challenges and complexities involved in it. Our research includes the development of a deep neural network model trained on extensive sign language datasets developed by us, enabling robust and real-time recognition of gesture.

II. Motivation

Sign language plays a crucial function in facilitating communication for people with speech impaired and hard-of hearing disabilities, offering them a means to express themselves, engage with others, and access information and services in their native languages. This compelling need for improved accessibility and inclusivity underscores the importance of our paper that aims to leverage the power of deep learning to develop an innovative solution. The main aim of this paper is to address the communication challenges encountered by individuals with impairments in listening and speaking. By developing animated sign language, we aim to offer a more accessible and natural method for conveying thoughts, emotions, and needs effectively. Moreover, our research endeavors to empower individuals with disabilities by granting them the ability to express themselves independently, thereby reducing their dependence on intermediaries or unavailable technology. Leveraging the capabilities of deep learning, we aim to create a precise system for recognizing and generating sign language gestures, thereby enhancing communication efficiency and intuitiveness. In alignment with the broader societal goal of inclusivity, our paper strives to foster a world where everyone can participate fully by providing individuals with listening and speaking disabilities with the means to communicate effectively, ultimately contributing to a more inclusive and equitable society.

III. Outline Of The Paper

The "Animated Sign Language" system is a novel solution designed to overcome communication challenges for deaf and dumb. It uses deep learning techniques to translate sign language into written or spoken language in real-time. The system is built on a dataset of sign language gestures and uses Recurrent Neural Networks (RNNs) for gesture recognition. NLP is employed to enhance the fluency of the translated text.

IV. Literature Survey

Matten Ahmed et al. [1] introduced 'Deaf Talk,' a Kinect V2-based 3D Animated Sign Language Interpreter. Their aim was to facilitate communication between deaf individuals and those who can hear, with the objective of overcoming challenges related to speech and hearing impairments.

Hamzah Luqman et al. [2] introduced a highly effective two-stream network designed for the recognition of isolated sign language. Their approach involved utilizing accumulative video motion for enhanced recognition accuracy.

Zhibo Wang et al. [3] introduced a system for identifying hand gestures. This system addresses the communication barrier between the listening-impaired community and the general population. Unlike previous sign language recognition systems, it provides continuous recognition with enhanced accuracy. It effectively addresses challenges associated with sign segmentation and accurately captures both finger and arm motions during sign language interpretation.

Karly Kudrinko et al. [4] created an extensive and thorough review that focused on the importance of hand gestures as a primary mode of contact for non-verbal individuals and deaf. The review outlined the communication hurdles experienced by these populations when interacting with individuals unfamiliar with sign language.

Mohammed Algabri et al. [5] introduces a novel system for recognizing Sign Language in Arabic. This system utilizes specific 2D body and hand key points extracted from consecutive video frames. This is achieved by integrating a 3D recurrent neural network (RNN) skeleton network with a 2D point convolution network.

Jing Yu et al. [6] focus on automating detection of traffic signs and in advanced transportation systems, crucial for advancing autonomous driving. Unlike conventional approaches analyzing single images, their fusion model combines YOLO-V3 and VGG19 networks to exploit multi-image relationships for swift and precise traffic sign detection in video sequences.

Pooja Chaudhari et al. [7] approach involves creating Indian Sign Language (ISL) gestures from various inputs such as text, subtitles from video content, audio, and text within images. For our study, we specifically analyze You tube videos. Through additional processing and the utilization of SiGML files, we generate animated gestures as the final output. Additionally, we propose a method for recognizing ISL gestures by employing Recurrent Neural Networks (RNNs) are utilized to extract both spatial and temporal features.

Sathiamoorthy Manoharan et al. [8] introduces an original method, inspired by previous studies, aimed at overcoming the unidirectional communication obstacle with a focus on accessibility and precision. It ensures accessibility through mobile device utilization for user interactions while ensuring accuracy by employing a transformer model. The system operates via a four-stage process facilitating one-way communication between individuals with hearing impairments and those without: audio capture, conversion of audio to text, transliteration of text to gloss, and animation of gloss.

V. Limitations And Challenges

Despite its innovative approach, the "Animated Sign Language" system faces several limitations and challenges. Firstly, the accuracy of sign language recognition, even with RNNs, may be influenced by the diversity of sign languages and individual variations, potentially leading to misinterpretations. The use of 3D gestures, while engaging, might not suit all users' preferences and could pose usability challenges. Furthermore, customization of gestures may require technical skills that some users lack. Lastly, ensuring widespread adoption and accessibility, especially in underserved communities, may require addressing issues such as cost, technological infrastructure, and user training.

VI. System Design

The system begins with an Input Module that captures user sign language gestures, employing a camera or similar gesture recognition devices. These captured gestures then undergo a Preprocessing Module aimed at noise reduction and preparing them for subsequent processing. Techniques like normalization, segmentation, and feature extraction are applied to refine the gestures. Subsequently, these pre-processed gestures are fed into a Deep Learning Model specialized in recognizing sign language gestures, selected based on gesture complexity. The output from this model undergoes a Post-processing Module, converting it into a format suitable for generating animations.

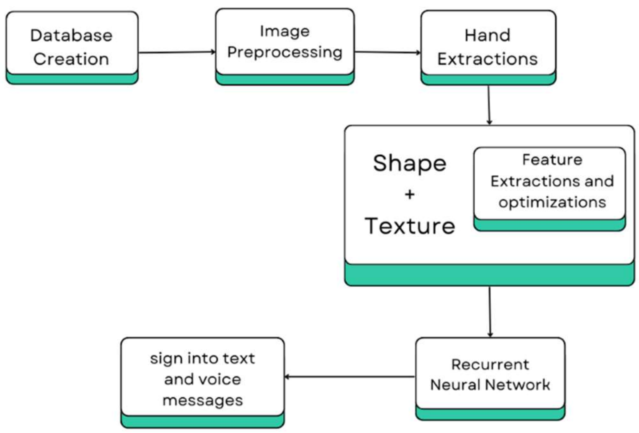

Fig. 1. Sign to text module

This stage involves mapping recognized gestures to corresponding animations. The Animation Generation Module then utilizes this processed data to create animations representing the recognized gestures, which are finally displayed to the user. Additionally, a crucial component is the Feedback Loop, allowing the system to learn from errors and refine its recognition capabilities over time, ensuring continuous improvement in gesture recognition accuracy.

In the first module, we specially employ Recurrent Neural Networks, to convert sign gestures performed by individuals with hearing and speaking disabilities into text and subsequently into audio. The input is acquired through a camera or another input device, and the image processing stage enhances the features of the sign gesture. Using a pre-trained RNN model, we recognize the sign gesture, translating it into text through a sign language to text conversion algorithm. This text will subsequently be inputted into a Text-to-Speech (TTS) engine converting it into clear and understandable audio. The entire process is intended to offer a smooth and efficient methods for individuals with hearing and speaking disabilities to comprehend and communicate using a combination of sign language interpretation and auditory feedback.

In the second module, our focus shifts to catering specifically to the deaf community by converting audio input into animations. We utilize advanced techniques, including deep learning and animation generation algorithms, to transform spoken words into visual representations.

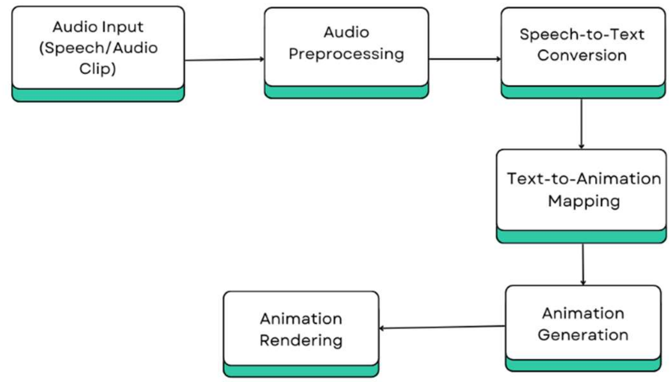

Fig. 2. Speech to sign module

The audio signals are processed and analysed to identify the corresponding movements and expressions associated with the spoken words. This information is then used to generate animated characters or gestures that mimic the gestures, ensuring that deaf individuals can comprehend spoken language through visually intuitive animations. This module seeks to narrow the communication divide between individuals with listening impairments and those without, providing a more inclusive and comprehensive communication experience.

VII. EXPERIMENTAL RESULTS

The key objective is to create a system leveraging deep learning techniques that facilitate communication for individuals with hard-of-hearing and speaking disabilities. This involves developing a system capable of converting sign language into audio and speech into animation. The initial phase involves comprehensive research on sign language gestures, suitable deep learning models for gesture recognition, and animation methods. Gathering requirements from stakeholders like users, caregivers, and medical professionals will be crucial in shaping the system's functionalities. A hypothesis has been formulated, postulating that a deep learning model can accurately recognize sign language gestures and convert them into audio, thereby enhancing communication for disabled individuals. The experimental design entails trained model with a sign language gesture's dataset and evaluating its performance on a distinct test set.

The system's effectiveness will be measured based on its accuracy in recognizing gestures and translating them into animations. A prototype system will be developed, comprising modules for gesture capture, preprocessing, deep learning-based gesture recognition, post-processing, and animation generation. Subsequently, rigorous testing will be conducted with users fluent in sign language, collecting data on the system's performance in recognizing gestures and translating them into animations.

Analysis of the gathered data will be conducted to assess the model's accuracy in recognizing gestures and the system's effectiveness in translating them into animations. Results obtained will be interpreted to determine if the system fulfills its objectives and supports the initial hypothesis. Based on these outcomes, conclusions will be drawn regarding the system's effectiveness, accompanied by recommendations for further enhancements. A comprehensive report detailing the entire experimental process, from research and requirement gathering to prototype development, testing, data analysis, conclusion drawing, and recommendation making, will be documented. Feedback from users and stakeholders will be incorporated into subsequent iterations to continually improve the system's effectiveness. Furthermore, the project's findings and insights will be communicated to relevant stakeholders, team members, and interested parties through presentations and effective communication channels.

VIII. Workproof And Diagrams

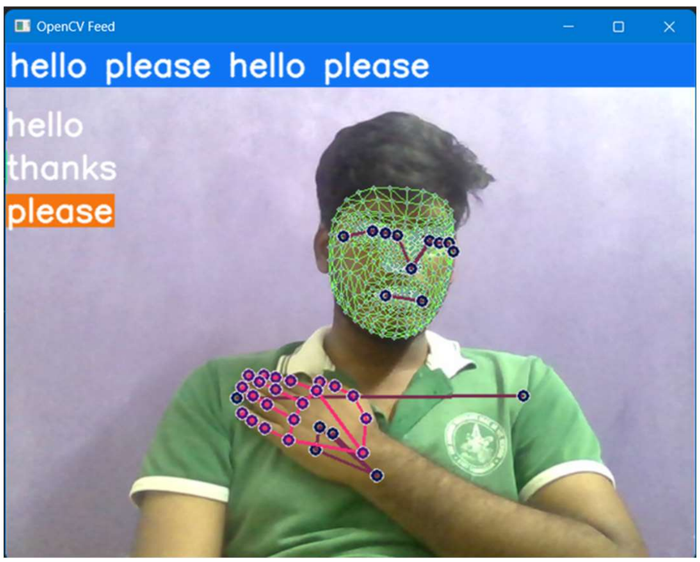

Fig. 3. Detection of the Sign "Please"

Fig. 4. Detection of the Sign "Hello"

In this we present an AI model try to achieves above 90% accuracy in detecting both the face and hands of users. Our model utilizes key modules such as MediaPipe for robust performance. The paper includes photographic evidence from our experimentation phase, showcasing the successful implementation and validation of our AI system's capabilities. These images serve as proof of the model's accuracy and effectiveness in facilitating animated sign language communication for individuals with hearing and speaking disabilities.

IX. Conclusion

In conclusion, the "Animated Sign Language" system is a phenomenal technological breakthrough that offers an unparalleled solution to the communication hurdles faced by deaf and mute individuals. Its core strength lies in its seamless translation of a wide array of sign language gestures into written or spoken language in real-time. This functionality empowers users to communicate effectively with a broader audience, transcending language barriers and promoting inclusivity in everyday interactions. By incorporating natural language processing, the system also enhances the fluency and coherence of the translated text, facilitating more meaningful and comprehensible communication.

Moreover, the introduction of customizable 3D gestures is a groundbreaking feature that elevates the communication experience to new heights. These gestures not only visualize sign language expressions but also inject a level of engagement and personalization crucial for fostering genuine connections. The "Animated Sign Language" system's demonstrated high accuracy in gesture recognition and translation makes it a potent tool for improving the lives of individuals with hearing and speech disabilities, promoting inclusivity, and breaking down communication barriers in a rapidly evolving world. It represents a remarkable step forward in bridging the divide and fostering a more empathetic and accessible society.

References

[1] M. Ahmed, M. Idrees, Z ul Abideen, R. Mumtaz, Sana Khalique, "Deaf talk Using 3D Animated Sign Language: A Sign Language Interpreter Using Microsoft's Kinect V2" in SAI Computing Conference, Sep. 01 2016, doi:10.1109/SAI.2016.7556002.

[2] H. Luqman, "An efficient two stream network for isolated sign language recognition using accumulate video motion", IEEE Access vol. 10, pp. 93785 - 93798, Sep.2022, doi:10.1109/ACCESS.2022.3204110.

[3] Zhibo Wang et al, "Hear Sign Language: A Real-time End-to-end Sign Language Recognition System", IEEE Transactions on Mobile Computing vol. 21, pp. 2398 - 2410, Jul. 01 2022, doi:10.1109/TMC.2020.3038303.

[4] K. Kudrinko, E. Flavin, X. Zhu, Qingguo Li. "Wearable Sensor-Based Sign Language Recognition: A Comprehensive Review". IEEE Reviews in Biomedical Engineering. vol. 14, pp. 82 - 97, Aug. 26 2020, doi:10.1109/RBME.2020.3019769.

[5] M. Algabri et al, "Arabic Sign Language Recognition System Using 2D Hands and Body Skeleton Data" in IEEE Access, vol. 9, pp. 59612 – 59627, Mar 30 2021, doi:10.1109/ACCESS.2021.3069714.

[6] Jing Yu, Xiaojun Ye, Qiang Tu, "Traffic Sign Detection and Recognition in Multi-images Using a Fusion Model with YOLO and VGG Network" in IEEE Transactions on Intelligent Transportation Systems, Vol. 23, Sep 2022, doi: 10.1109/TITS.2022.3170354.

[7] Pooja Chaudhari, Mangesh Bedekar, “Indian Sign Language Generation – A Multi-modal Approach” in 14th International Conference on Computing Communication and Networking Technologies (ICCCNT),Jul 8 2023,doi: 10.1109/ICCCNT56998.2023.10306461.

[8] Xinfeng Ye, Zhongling Tang, Sathiyamoorthy Manoharan,“From Audio To Animated Signs” in IEEE-9th International Conference on Electrical and Electronics Engineering (ICEEE). 16 May 2022, doi:10.1109/ICEEE55327.2022.9772564.