Review Of Sign Language Recognition And Translation To English And Marathi

Pratik Dahatonde
Department of Computer Engineering
Nutan Maharashtra Institute of Engineering and Technology
Pune

Prathamesh Khandekar
Department of Computer Engineering
Nutan Maharashtra Institute of Engineering and Technology
Pune

Omkar Kharat
Department of Computer Engineering
Nutan Maharashtra Institute of Engineering and Technology
Pune

Dr. Saurabh Saoji
Department of Computer Engineering
Nutan Maharashtra Institute of Engineering and Technology
Pune

Dr. Naveenkumar Jayakumar
Department of Computational Intelligence
Vellore Institute of Technology
Vellore

Abstract—Sign Language Recognition and Translation to English and Marathi is a project that aims at enhancing communication between the deaf and the hard of hearing (DHH) community with the hearing populace. Sign language, which is the primary means of communication among DHH, is often a barrier to various aspects such as education, health care, employment and social interactions. This project seeks to eliminate this communication barrier using state-of-the-art technology by recognizing sign-language gestures for spoken and written language in English and Marathi—two majorly spoken languages. The project serves an essential purpose of improving communication between the Deaf-Hard of Hearing (DHH) community and the general population. A primary form of communication used by DHH is sign language, which remains problematic in different areas such as education, health care, jobs or even social interaction. In order to solve this problem, cutting-edge technologies are utilized in order to create an innovative solution which can detect sign language gestures and translate them into spoken or written forms in both English as well as one of the major Indian languages – Marathi.

Keywords–Machine Learning, MediaPipe, OpenCV, LSTM Neural Network, Sign Language

I. Introduction

The project is a signal of how far advanced technology has gone in dealing with communication challenges facing the Deaf and Hard of Hearing (DHH) community. Sign language is an important mode of communication for the DHH but it often presents barriers in educational, healthcare, employment and social settings. This work is a breakthrough attempt to bridge this gap through employment of cutting-edge technologies such as Long Short-Term Memory (LSTM) networks for sign language recognition and MediaPipe for accurate hand and gesture tracking. The target is to convert sign language gestures fluently into spoken or written English and Marathi giving new ways of communicating with each other. What makes this project important is not just its technological superiority but its ability to enable the DHH community become inclusive across diverse domains. With all details on system design intricacies, data processing algorithms, real time capabilities, we are guided by an overarching vision that speaks to breaking down barriers to communication with our society moving towards becoming more interconnected as well as inclusive.

II. Literature Survey

Sign language is a language used to communicate with people who are deaf and have speech difficulties. When the user's sign language interferes with the user's non-sign language, it may be difficult for the user's sign language to express itself to others. Language recognition systems can help signers interpret sign language other than the user's language. The study demonstrates the language recognition ability of the Arabic language from RGB video recordings. To achieve this goal, two datasets are considered, namely the original data and the facial segmentation dataset created from the original data. Additionally, the study presents ”SelfMLP”, a multilayer perceptron, as a functional layer to develop a CNN-LSTM-SelfMLP model for Arabic language recognition. CNN and three SelfMLPs based on MobileNetV2 and ResNet18 were used to create six different CNN-LSTM-SelfMLP architecture models for the comparative performance of Arabic language recognition. The study examines a simple signature model for handling emergency events. [1]

In the paper it is discussed how the image can be processed by software to create a logo in real time. The next stage is to use classifiers to differentiate between individual characters and output translated text. Various Machine Learning techniques can be used to train the model using the data. Using efficient algorithms and best-in-class data and advanced hardware, the device seeks to improve existing systems in terms of response time and accuracy within its field. Current systems are only able to identify movement through image processing with some delay. With this project, the research aims at constructing an intuitive, reliable system for people with speech or hearing troubles that is usability in day-to-day situations. [2]

The communication gap between normal people and deaf-mute is responsible for their separation from society. Today, people can capture any gesture made by a deaf-mute individual thanks to technological advancement and turn it into text using machine learning. In the work, a comprehensive model is used for recognizing gestures that uses MediaPipe to extract hand and gesture positions and short-term memory (LSTM) to express and recognize gestures. It will convert Bangla sign language gestures into easily readable text. Our proposed model for this research was based on the requirements analysis which will be conducted in four phases; data collection and pre-processing; selecting an appropriate neural network structure, training it with the available data and then testing it; real-time experimentation. The self-created dataset of Bangla sign language helps teach the model to recognize gestures. To make sure that system functions optimally a new way of helping humans communicate with one another even when there is no common spoken language between them was developed.[3]

The paper provides a comprehensive overview of the state-of-the-art techniques and advancements in sign language recognition using deep learning methods. The authors, Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera, delve into the challenges and opportunities in this field, highlighting the importance of sign language recognition for facilitating communication and accessibility for the hearing-impaired community. The survey covers various aspects of sign language recognition, including dataset collection, preprocessing techniques, feature extraction methods, and different deep learning architectures employed for this task. The authors discuss the significance of hand gesture recognition, facial expression analysis, and body pose estimation in enhancing the accuracy of sign language recognition systems. Furthermore, the paper explores the applications of sign language recognition in real-world scenarios such as human-computer interaction, assistive technologies, and educational tools. The authors also address the limitations and future research directions in the field, emphasizing the need for robust and efficient algorithms to improve the performance of sign language recognition systems. Overall, the paper provides a valuable insight into the advancements and challenges in sign language recognition using deep learning techniques, highlighting its potential impact on improving accessibility and communication for the deaf and hard of hearing community.[4]

In the research paper, the authors propose a novel approach to dynamic hand gesture recognition using a two-layer Bidirectional Recurrent Neural Network (BRNN) with a Leap Motion Controller (LMC). They address the need for improved accuracy in dynamic hand gesture recognition, crucial for human-computer interaction. The system is tested on American Sign Language (ASL) datasets and Handicraft-Gesture datasets, achieving high accuracies. The proposed system demonstrates superior or comparable performance compared to existing literature. Key contributions include accurately determining gesture start and end times, selecting discriminative features based on fingers and palm, and introducing a two-layer BRNN for improved recognition accuracy. This work pioneers the combination of BRNN with LMC features, setting a new standard for dynamic gesture recognition.[5]

The research is driven by the urgent need to enable more than 200,000 hearing and speech-impaired individuals in Bangladesh to integrate into the mainstream workforce through the development of a real-time Bangla Sign Language (BdSL) interpreter. The recognition of BdSL presents complex challenges, including variations in skin tone, hand orientation, and background settings, necessitating sophisticated computer vision and deep learning approaches. To tackle these challenges head-on, the study undertakes the ambitious task of leveraging deep machine learning models on meticulously curated and robust datasets. A significant milestone achieved in this endeavor is the creation of a comprehensive BdSL dataset, meticulously designed to encompass a wide spectrum of backgrounds, skin tones, and hand representations. This database serves as the cornerstone for reducing inter-class similarity and enhancing recognition accuracy. Building upon this foundation, the research proposes innovative BdSL recognition systems by harnessing the power of transfer learning on three distinct pre-trained deep convolutional neural networks (CNNs). Moreover, the study pioneers the integration of semantic segmentation-based hand detection techniques, offering a refined approach for precisely identifying hand gestures before initiating the recognition process. This methodology, coupled with transfer learning on pre-trained CNNs, results in the development of robust recognition systems that outperform existing benchmarks. Notably, the best-performing model achieves an exceptional accuracy rate of 99.99%, showcasing the efficacy of the proposed approach. The culmination of this research effort is the creation of a real-time BdSL interpreter, poised to revolutionize communication accessibility for hearing and speech-impaired individuals in Bangladesh. By providing them with effective communication tools, this technological innovation has the potential to foster inclusivity and empowerment, thereby breaking down barriers and promoting social integration. The paper, structured with meticulous attention to detail, guides readers through the comprehensive research process, from the contextualization of the problem to the presentation of groundbreaking findings and recommendations for future exploration and development in the field of BdSL interpretation technology.[6]

The paper introduces the significance of analyzing time-series data for identifying long-term trends and making predictions, highlighting the effectiveness of Long Short-Term Memory (LSTM) recurrent neural networks (RNNs) in handling time-series forecasting problems. LSTM's gated structure and flexible network topology make it an attractive tool over conventional methods. The paper proposes a hardware implementation of LSTM network architecture for time-series forecasting, utilizing TSMC 0.18 m CMOS technology and HP memristor model. The introduction provides a brief history of LSTM and its advantages, followed by a comparison with previous work and the proposal of voltage-based circuits for higher accuracy. The paper aims to address time-series prediction challenges through analog hardware implementation, structured with problem description, LSTM overview, circuit introductions, simulation results, and conclusions.[7]

The paper addresses the significance of hand gesture recognition, particularly in sign language translation, and outlines the challenges in developing an efficient recognition system, such as hand segmentation, local and global feature representation, and gesture sequence modeling. It proposes a novel system for dynamic hand gesture recognition using multiple deep learning architectures, evaluated on a challenging dataset. The introduction emphasizes the importance of hand gesture recognition in human-computer interaction applications and sign language translation. Previous methods are discussed, highlighting the limitations of existing systems in capturing the complexity of sign language gestures. The paper then explores the application of deep learning techniques, including convolutional neural networks (CNNs) and 3D CNNs, for hand gesture recognition, addressing drawbacks such as temporal dependence modeling and feature relevance. The contributions of the paper include optimizing knowledge transfer in the C3D architecture, presenting a hand gesture recognition system based on optimized architectures, introducing a novel hand segmentation method, and optimizing architectures for local features aggregation. The structure of the paper is outlined, covering related works, dataset description, proposed system, experimental results, and conclusions.[8]

The paper addresses the challenge of sign language recognition using computational models, particularly focusing on isolated sign language recognition (SLR). It introduces Motion History Images (MHI) generated from RGB video frames as a representation of spatio-temporal information in sign videos. The RGB-MHI images effectively summarize each sign video in a single RGB image. Two approaches are proposed using the RGB-MHI model: integrating it as a motion-based spatial attention module into a 3D-CNN architecture and using RGB-MHI model features directly with a late fusion technique. Extensive experiments are conducted on large-scale isolated sign language datasets, demonstrating that the proposed models using only RGB data can compete with state-of-the-art models in the literature that use multi-modal data. The introduction outlines the importance of sign language recognition for deaf communities and the growth of research in this area. It distinguishes between isolated and continuous SLR and highlights the challenges specific to isolated SLR, such as discriminating fine-grained local motions and variations in hand gestures, facial expressions, and trajectories. The proposed solution aims to address these challenges using only RGB images without explicit part segmentation or additional data modalities.[9]

The paper addresses the communication barriers faced by individuals with impaired speaking or hearing abilities and proposes a deep learning-based model for detecting and recognizing words from a person's gestures in Indian Sign Language (ISL). Using LSTM and GRU models, the paper recognizes signs from isolated ISL video frames, achieving around 97% accuracy over 11 different signs. The proposed system aims to bridge the communication gap between individuals who know sign language and those who don't, thus providing equal communication opportunities. It outlines the prevalence of speech and hearing impairments in India and the importance of sign language as a visual communication mode. The system processes videos of hand gestures, predicts words, and generates meaningful sentences, facilitating communication. The study focuses on static and dynamic signs, developing a dataset (IISL2020) comprising 11 commonly used words and 630 samples. Unlike other sign language datasets, IISL2020 was created without external aids like sensors or smart gloves, considering natural conditions.[10]

III. PROPOSED SYSTEM

The proposed system will use image processing techniques to recognize signs from a video sequence and classify them into their respective categories. System will use a combination of Holistic Pipeline and Long Short-term Memory (LSTM) network in order to extract features from the images and recognize the signs.

1. Data collection: In the data collection phase, we collect individual gestures. We will use one index to label all gestures in order to prepare the data for training and testing –the first gesture will have 0 as its label, the second one 1, and so on.

2. Data preparation: Using various data preparation techniques to format the main content extracted from the data. In the data model, all values of each order are stored

as a NumPy array (X), which then corresponds to another NumPy array of tags (Y).

3. Implementation Procedure: A dataset was created that would ensure accurate landmark identification of user's hands and body parts using MediaPipe solution by Google and Python's OpenCV library.

4. Model Generation and Model Testing: Once the model has been created, it will be reviewed and improved to ensure it performs accordingly and maintains a certain accuracy.

5. User Interface: The proposed system will have user-friendly interfaces for Deaf and Hard of Hearing (DHH) users as well as users with normal hearing. We prioritize accessibility.

IV. System Design and Architecture

A. System Design

Sign Language Recognition and Translation to English and Marathi project's system design consists of an architecture that is well-structured with several major components. These include the LSTM model for sign language recognition that has been integrated with MediaPipe for hand and gesture tracking. Inputting sign language becomes much easier through a user-friendly interface while data preprocessing methods improve the quality of input data. Such real time processing possibilities are facilitated by continuous learning mechanisms which ensure adaptability and improvement over time. Measures have been taken to protect the privacy of users during communication. Among other things, it includes scalability, error and performance metrics that show its design, efficiency, user satisfaction and make it more interesting for others. The overall technology process, process requirements, and verification process are also discussed to ensure a complete understanding of the system under development.

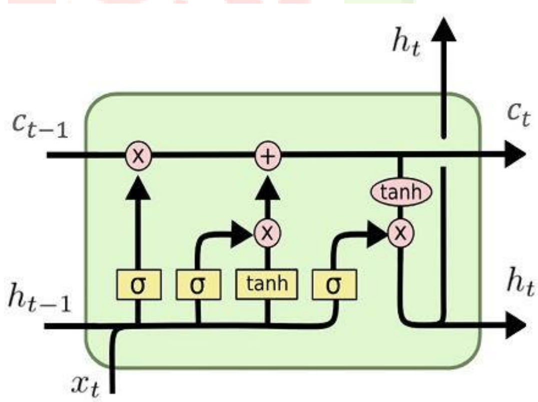

Fig. 1. LSTM Cell Model.

B. Algorithm

Step 1: Get started

Step 2: Collect a diverse dataset of sign language gestures in both English and Marathi.

Step 3: Prepare the data for training and testing, we will label all the gestures with a single index, where the first gesture will be labeled with 0, the second one will be 1, and so on.

Step 4: A dataset construction process is performed and landmarks are extracted in the dataset module.

Step 5: The data is then processed according to the specifications required for the model and sent to the LSTM Module for analysis.

Step 6: Instantaneous real-time feedback is generated by the camera and analyzed by the model after adjusting the hyperparameters.

Step 7: The results are displayed as text next to the appropriate gesture.

C. System Architecture

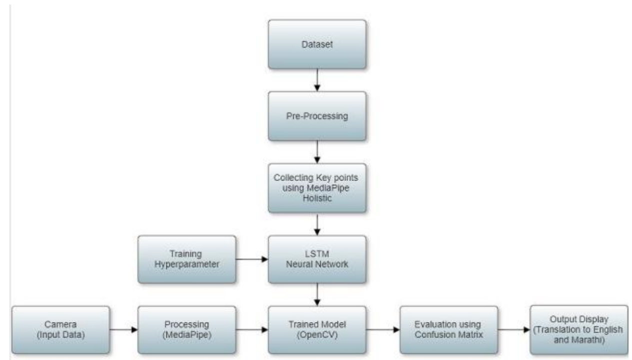

Fig. 2. System Architecture

V. Results and Discussion

It's important to note that the results and discussions presented herein reflect the outcomes of the initial phase of our comprehensive research project on "Sign Language Recognition and Translation to English and Marathi." In this phase, the integration of MediaPipe Holistic has been a focal point, and the outcomes are promising, serving as a foundational element for the broader system development. The successful implementation of MediaPipe Holistic underscores its potential in capturing and interpreting complex sign language gestures, particularly in real-time scenarios. Figure 3 and Figure 4 provide visual representations of the working of the MediaPipe Holistic detection, demonstrating its ability to accurately track and interpret sign language gestures. However, it is crucial to emphasize that this represents only the initial segment of our multifaceted project.

As the project moves forward, subsequent phases will build upon these results, incorporating additional technologies and refining the system for improved accuracy and inclusivity. The insights gained from this initial integration will guide the continued development of our sign language recognition system, ensuring that it meets the diverse and evolving needs of the Deaf and Hard of Hearing community. The discussion within this section serves as a springboard for the holistic

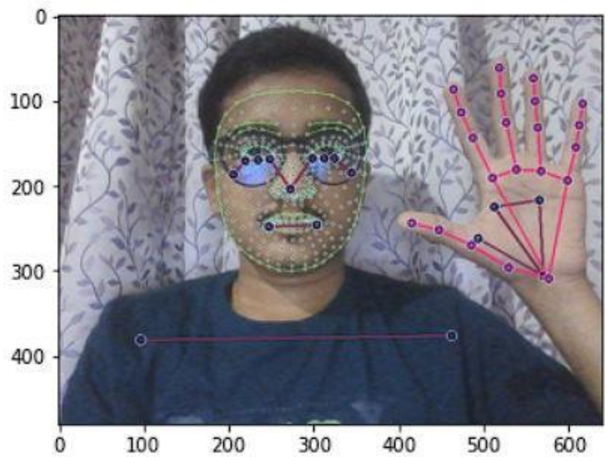

Fig. 3. MediaPipe Holistic Hand tracking System (Left Hand- Number 5)

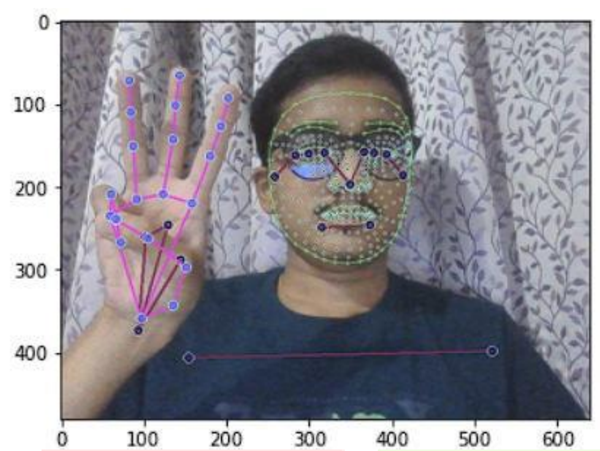

Fig. 4. MediaPipe Holistic Hand tracking System (Right Hand- Number 3)

exploration of our project, highlighting the accomplishments thus far while setting the stage for the broader advancements and contributions to come.

VI. Conclusion

In conclusion, the successful integration of MediaPipe Holistic marks a significant milestone in the initial phase of our research project on ”Sign Language Recognition and Translation to English and Marathi.” The outcomes highlight the robust capabilities of MediaPipe Holistic in accurately capturing and interpreting intricate sign language gestures, laying a solid foundation for the broader system development. While this phase focuses on the implementation of MediaPipe Holistic, it serves as a promising precursor to the comprehensive advancements planned for the future. The insights gained and challenges addressed during this phase will inform subsequent stages, contributing to the refinement and optimization of our sign language recognition system. As we progress, the integration of additional technologies and the continuous fine-tuning of the system will enhance its accuracy, usability, and inclusivity. This initial success underscores our commitment to creating an advanced, inclusive, and impactful solution for the Deaf and Hard of Hearing community. The journey has just begun, and the outcomes from this phase propel us forward with enthusiasm toward the broader goals and contributions envisaged in the subsequent stages of our research.

Acknowledgment

The authors wish to thank their parents and associates for their valuable support and encouragement throughout the development of the project work, and they would also like to thank their guide Dr. Saurabh Saoji for guiding them throughout the project work.

References

[1] Podder KK, Ezeddin M, Chowdhury MEH, Sumon MSI, Tahir AM, Ayari MA, Dutta P, Khandakar A, Mahbub ZB, Kadir MA. Signer-Independent Arabic Sign Language Recognition System Using Deep Learning Model. Sensors (Basel). 2023 Aug 14;23(16):7156. doi: 10.3390/s23167156. PMID: 37631693; PMCID: PMC10459624.

[2] G. M. Rao, C. Sowmya, D. Mamatha, P. A. Sujasri, S. Anitha and R. Alivela, ”Sign Language Recognition using LSTM and Media Pipe,” 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 1086-1091.

[3] M. W. Foysol, S. E. A. Sajal and M. J. Alam, ”Vision-based Real Time Bangla Sign Language Recognition System Using MediaPipe Holistic and LSTM,” 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 2023, pp. 19-24.

[4] Razieh Rastgoo, Kourosh Kiani, Sergio Escalera, Sign Language Recognition: A Deep Survey, Expert Systems with Applications, Volume 164, 2021, 113794, ISSN 0957-4174.

[5] Yang, Linchu Chen, Ji’An Zhu, Weihang. (2020). Dynamic Hand Gesture Recognition Based on a Leap Motion Controller and Two-Layer Bidirectional Recurrent Neural Network. Sensors. 20. 2106. 10.3390/s20072106.

[6] Podder, K.K.; Chowdhury, M.E.H.; Tahir, A.M.; Mahbub, Z.B.; Khandakar, A.; Hossain, M.S.; Kadir, M.A. Bangla Sign Language (BdSL) Alphabets and Numerals Classification Using a Deep Learning Model. Sensors 2022, 22, 574.

[7] Adam, Kazybek Smagulova, Kamilya James, A.. (2018). Memristive LSTM network hardware architecture for time-series predictive modeling problems. 459-462. 10.1109/APCCAS.2018.8605649.

[8] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M.A. Bencherif, T.S. Alrayes, et al., ”Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation”, IEEE Access, vol. 8, 2020, pp. 192527-192542.

[9] Ozge Mercanoglu Sincan, Hacer Yalim Keles, ”Using Motion History Images With 3D Convolutional Networks in Isolated Sign Language Recognition”, IEEE Access, vol.10, 2022, pp.18608-18618.

[10] Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil González, A.B. and Corchado, J.M., 2022. Deepsign: Sign language detection and recognition using deep learning. Electronics, 11(11), p.1780

[11] Thasin Abedin, Khondokar SS Prottoy, Ayana Moshruba, and SafayatBin Hakim. Bangla sign language recognition using concatenated bdslnet-work. arXiv preprint arXiv:2107.11818, 2021.

[12] Abey Abraham and V Rohini. Real time conversion of sign languageto speech and prediction of gestures using artificial neural network.Procedia computer science, 143:587–594, 2018.

[13] A Sachin Agrawal, Agnishrota Chakraborty, and M Rajalakshmi. Real-time hand gesture recognition system using mediapipe and lstm. Journalhomepage: www. ijrpr. com ISSN , 2582:7421, 2022.

[14] Md Alam, Mahib Tanvir, Dip Kumar Saha, Sajal K Das, et al. Two di-mensional convolutional neural network approach for real-time ban-glasignlanguage characters recognition and translation. SN Computer Sci-ence, 2(5):1–13, 2021.

[15] Najeefa Nikhat Choudhury, Golam Kayas, et al. Automatic recognitionof Bangla sign language. PhD thesis, BRAC University, 2012.