Translation of Indian Sign Language to Text using an Edge Device

Keshava.A[1], Murugan.M[2], Prakash.S[3] and Malathi.S[4]

[1-3] UG Students, [4]Assistant Professor, Department of Electronics and Communication Engineering, National Engineering College, Kovilpatti, India

E-mail: 2111113@nec.edu.in, 2111404@nec.edu.in, 2111110@nec.edu.in, malathi_ece@nec.edu.in

Abstract

This project is focused on developing an efficient real-time system for sign language recognition that can be translated into textual output, using Raspberry Pi integrated with an OLED display. Therefore, this will enable communication across all angles, for those who are deaf or mute, to effectively and readily communicate among people with any other normalcy. The system will utilize an optimized lightweight deep model suitable for edge devices that can detect and translate hand gestures representing letters within the Indian Sign Language alphabet. The system uses Media Pipe framework in hand gesture recognition to locate hand landmarks. These landmarks are given as input into a neural network to classify the gestures into corresponding letters. The final output is displayed through an OLED screen, so the system can be both taken out of the laboratory and used in everyday life. It ensures that the computational cost and power consumption will be quite optimal to deploy even in resource-constrained devices like Raspberry Pi through the use of pre-trained models such as Mobile Net or Squeeze Net. It is socially designed, aimed to enable the accessibility of the communication among the deaf and mute community, and also provide an easy, cost-effective solution to assist daily interaction. This project further extends further possible improvements: for example, increasing classes in gestures, as well as immediate feedback.

Keywords Real time system, Raspberry pi, Gesture to text, accessibility.

1 INTRODUCTION

Sign language recognition may be considered an emergent technology, as it forms a critical means to fill the gap in communication that the people with hearing or speech impairments face. The new wave of advancements in deep learning, which mainly revolved around the design of CNNs, has really revolutionized the real-time recognition an interpretation of sign language gestures. In order to create an inclusive environment that facilitates proper communication with others in everyday situations, it is very important to recognize hand gestures speedily and accurately.

- Capture Hand Gestures: Use a camera module connected to the Raspberry Pi Zero to capture live hand gestures.
- Hand Detection and Landmark Identification: Employ MediaPipe to detect hands in the camera feed
- and extract landmarks from the detected hands, such as the position and movement of fingers.
- Gesture Classification: Train a custom deep learning model capable of recognizing 36 ISL alphabets (A-Z, and 0-9) from the detected hand landmarks.
- Output Text Display: Display the recognized gestures on a small OLED screen, allowing for real-time communication.

[1] deals with the issue of communication between hearing-impaired people by identifying signs for emergencies in Indian Sign Language (ISL) using deep learning models. An application-specific dataset in the form of video frames capturing eight different emergencies was utilized, and the frames were labeled to improve accuracy. [4] discusses an on-the-spot Indian Sign Language (ISL) numeral-to-text translation system that can enable easier communication among deaf people. Unlike traditional methods that rely on handcrafted features, the approach uses deep learning with a convolutional neural network (CNN) implemented in Keras. [11] addresses the communication gap between the deaf and mute community and the hearing population, who often lack understanding of sign language. It presents a comparative analysis of gesture recognition techniques using convolutional neural networks (CNNs) and machine learning algorithms.

The research compares three models: a pre-trained fine-tuned VGG16, a VGG16 with transfer learning, and a hierarchical neural network, all of which were trained using a self-designed dataset of Indian Sign Language (ISL) images of the 26 English alphabets. [8] aims to enhance the static sign recognition of Indian Sign Language (ISL) through deep learning and convolutional neural networks (CNN). It improves upon past attempts that tended to identify only a small number of hand signs and sought instead to adopt a more extensive methodology. A corpus of 35,000 images for 100 unique static signs was gathered from different users.

This project will create a real-time Indian Sign Language (ISL) recognition system, hence bringing gestures to text, utilizing a Raspberry Pi Zero integrated with an OLED display. The approach for building this system is carried out in several steps including data collection, model architecture design, model training, integration of the systems, and real-time testing on the edge device.

2 METHODOLOGY

2.1 Data Collection and Preprocessing

The methodology starts off with collecting the dataset for ISL hand gestures. For this project, a dataset of hand gesture images of ISL alphabet from A to Z and 1 to 9 is used. The images are captured through the camera module. Prior to processing or the pre-processing involves the dataset normalization of hand landmarks obtained from each gesture is carried out. The normalization is done by collecting the mediapipe output which is the hand landmarks, which are then stored in a csv file format. This would ensure the model generalizes well and is robust toward any gesture variation.

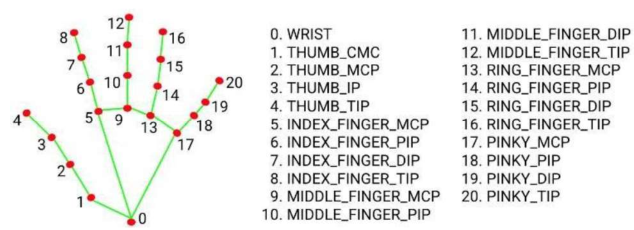

Figure 2.1 Hand Landmarks

For drawing hand landmarks for hand gesture recognition, MediaPipe is used for each image. MediaPipe is a powerful library developed by Google that detects and tracks 21 real time landmarks of the hand from a video stream. These features represent important points on the hand and are utilized as the basis of classification. Features so gathered are fed to a classification neural network, which classifies the gesture based on the patterns in the landmarks.

2.2 Model Architecture

The model contains a lightweight architecture and is optimized toward the edge device, the Raspberry Pi Zero. A CNN was chosen to process the hand gesture data stream. The architecture of the proposed model includes an input layer, convolutional layers, pooling layers, dense layers, and an output softmax layer. In CNN, it will detect a hand gesture from one of the predefined classes. As the computing constraint on edge devices is so strict, it optimizes the model on the edge devices by using lightweight pre-trained models such as MobileNet, which offer very high accuracy without very serious computational necessities.

2.3 Training the model

This model is then fitted to the processed dataset. The data is divided into training and test sets to evaluate the performance of the model on test data. For multi-class classification, a categorical cross-entropy loss function was utilized, and Adam was used as an optimizer since it efficiently handles learning rate adaptation. Validation accuracy monitored during training was also used to avoid over fitting. Early stopping in addition to model checkpoints with best performance was used at each training session, a model with the best performance was saved.

2.4 System Integration

The trained model is integrated with the Raspberry Pi and an OLED screen. The Raspberry Pi captures real- time video of hand gestures by attaching a camera module onto the device. It further processes the frames captured to deduce the hand landmarks using MediaPipe. These resultant landmarks are forwarded to the trained neural network for classification and the same is displayed in real time on the OLED screen. This integration will enable the system to identify gestures and respond immediately for successful communication.

2.5 Deployment and Testing

The last test involves the deployment and testing of the system for real-world conditions. This comprises the experimentation with the efficiency of real-time recognition of ISL gestures under different conditions, lighting, and backgrounds. The computationally efficiency of the deployed system is experimented on to see whether the system will run on the Raspberry Pi without lagging or consuming too much power. With this system, it has become suitable for accuracy, robustness, and ease of use to assist the deaf and mute people in their communications.

3 RESULTS

To test the trained model accuracy, the test dataset is used. With a batch size of 128, it is assumed to train for 1000 epochs with utmost precision on 99% validation set for all the letter and recognition: A – Z and 0 -
9. Total of 36 classes each with 600 datasets. Accuracy, precision, recall, and F1 score have been computed to verify whether the model is capable enough of recognizing individual classes, or gestures. The following table reflects metrics for each class:

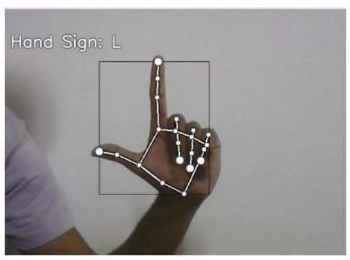

Figure 3.1 Hand Sign for L

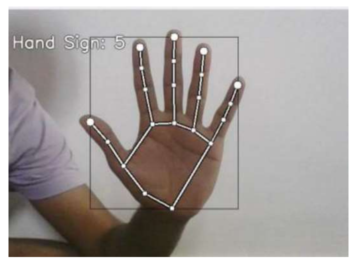

Figure 3.1 Hand Sign for 5

Label	Precision	Recall	F1-Score	Support
0	0.99	1.00	1.00	142
1	0.98	0.99	0.98	130
2	0.99	1.00	0.99	137
3	1.00	1.00	1.00	149
4	1.00	1.00	1.00	162
5	1.00	1.00	1.00	159
6	1.00	0.99	1.00	148
7	1.00	1.00	1.00	149
8	1.00	0.99	1.00	142
9	0.98	1.00	0.99	150
10	1.00	1.00	1.00	167
11	1.00	0.95	0.98	149
12	0.99	0.99	0.99	154
13	0.99	0.98	0.98	166
14	1.00	1.00	1.00	149
15	0.98	1.00	0.99	133
16	1.00	0.99	1.00	153
17	1.00	1.00	1.00	156
18	0.99	0.97	0.98	160
19	1.00	0.99	1.00	127
20	0.99	1.00	0.99	154
21	1.00	1.00	1.00	140
22	0.91	0.95	0.93	156
23	0.95	0.96	0.95	158
24	0.96	1.00	0.98	151
25	0.99	1.00	1.00	158
26	0.99	0.97	0.98	151
27	0.98	0.93	0.95	138
28	0.99	0.99	0.99	163
29	0.98	0.99	0.99	155
30	1.00	1.00	1.00	150
31	1.00	1.00	1.00	152
32	0.99	1.00	1.00	153
33	1.00	0.99	1.00	143
34	1.00	0.98	0.99	125
35	1.00	1.00	1.00	173

Initially, the project was implemented on the Raspberry Pi Zero due to its compact design and low power consumption. However, we encountered significant challenges related to the availability of essential software packages compatible with the ARMv6 architecture of the Raspberry Pi Zero. Several libraries and dependencies required for the Indian Sign Language recognition model, including those necessary for Mediapipe and the neural network framework, were not supported on this architecture.

To mitigate these challenges, we transitioned the implementation to the Raspberry Pi 4, which is based on the ARMv8 architecture. The Raspberry Pi 4 not only supports a wider range of packages and libraries but also offers enhanced processing capabilities, allowing for smoother execution of the model. This change enabled us to leverage the full potential of the software stack, ensuring better functionality and performance for real-time ISL recognition.

4 CONCLUSION

This project demonstrates the successful implementation of a real-time sign language recognition system using a lightweight deep learning model deployed on a Raspberry Pi , with the output displayed on an OLED screen. Even though there are many research in this domain, we distinguish our paper with the integration of OLED screen along with raspberry pi with an optimized neural network. By combining the MediaPipe framework for hand gesture detection and a neural network model optimized for edge devices, the system effectively translates simple Indian Sign Language gestures into text, allowing for seamless communication between the deaf and mute community and the general public. Ultimately, this project not only highlights the potential of edge computing for socially impactful applications but also provides a foundation for future research aimed at creating more advanced and inclusive systems for individuals with hearing and speech impairments. By continuing to optimize the system, we can ensure its practicality, affordability, and scalability, thus contributing to a more inclusive society.

5 FUTURE SCOPE

Despite the promising results, the system has its limitations, such as restricted gesture set and challenges posed by environmental variables like lighting and camera angles. However, it paves the way for future improvements that include expanding the gesture set to encompass the full Indian Sign Language words and sentences, Experimenting with advanced model architectures or transfer learning approaches, optimizing the model for better performance on resource-constrained devices, and exploring additional features like real-time feedback and multimodal inputs. These advancements would enhance the system's utility in real-world settings, making it a valuable tool for accessible communication.

6 REFERENCES

[1] Areeb, Q. M., Maryam, Nadeem, M., Alroobaea, R., & Anwer, F. (2022). Helping Hearing-Impaired in Emergency Situations: A Deep Learning-Based Ap- proach. IEEE Access, 10. https://doi.org/10.1109/ACCESS.2022.3142918

[2] Venugopalan, A., & Reghunadhan, R. (2021). Applying deep neural networks for the automatic recognition of sign language words: A communication aid to deaf agriculturists. Expert SystemswithApplications, 185. https://doi.org/10.1016/j.eswa.2021.115601

[3] Singh, A., Wadhawan, A., Rakhra, M., Mittal, U., Ahdal, A. al, & Jha, S. K. (2022). Indian Sign Language Recognition System for Dynamic Signs. 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions),ICRITO2022. https://doi.org/10.1109/ICRITO56286.2022.9964940

[4] Sajanraj, T. D., & Beena, M. v. (2018). Indian Sign Language Numeral Recognition Using Region of Interest Convolutional Neural Network. Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018. https://doi.org/10.1109/ICICCT.2018.8473141

[5] Bhagat, N. K., Vishnusai, Y., & Rathna, G. N. (2019). Indian Sign Language Gesture Recognition using Image Processing and Deep Learning. 2019 Digital Image Computing: Techniques and Applications, DICTA 2019. https://doi.org/10.1109/DICTA47822.2019.8945850

[6] Sruthi, C. J., & Lijiya, A. (2019). SigNet: A deep learning based indian sign language recognition system. Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019. https://doi.org/10.1109/ICCSP.2019.8698006

[7] Rao, G. A., Syamala, K., Kishore, P. V. V., & Sastry, A. S. C. S. (2018). Deep convolutional neural networks for sign language recognition. 2018 Conference on Signal Processing And Communication Engineering Systems, SPACES 2018, 2018-January. https://doi.org/10.1109/SPACES.2018.8316344

[8] Wadhawan, A., & Kumar, P. (2020). Deep learning-based sign language recognition system for static signs. Neural Computing and Applications, 32(12). https://doi.org/10.1007/s00521-019-04691-y

[9] Rathi, E., Luthra, J., & Sharma, A. (2021). Implementation and Performance Analysis of Indian Sign Language Recognition on various Computing Devices. Proceedings of the 2021 IEEE 18th India Council International Conference,
INDICON 2021. https://doi.org/10.1109/INDICON52576.2021.9691542

[10] Vidhyalakshmi, M. K., Jagadeesh, K., Dharmaraj, B. S., Sankar N, P., & Kumar, R.
G. (2024). Indian Sign Language Recognition using transfer learning with Efficient Net. Smart Science, 12(2). https://doi.org/10.1080/23080477.2024.2320956

[11] Sharma, A., Sharma, N., Saxena, Y., Singh, A., & Sadhya, D. (2021). Benchmarking deep neural network approaches for Indian Sign Language recognition. Neural Computing and Applications, 33(12). https://doi.org/10.1007/s00521-020-05448-8

[12] Subramanian, B., Olimov, B., Naik, S. M., Kim, S., Park, K. H., & Kim, J. (2022). An integrated mediapipe-optimized GRU model for Indian sign language recognition. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-15998-7

[13] Joshi, J. M., & Patel, D. U. (2024). Dynamic Indian Sign Language Recognition Based on Enhanced LSTM with Custom Attention Mechanism. SSRG International Journal of Electronics and Communication Engineering, 11(2). https://doi.org/10.14445/23488549/IJECE-V11I2P107

[14] Paul, A. J., Mohan, P., & Sehgal, S. (2020). Rethinking Generalization in American Sign Language Prediction for Edge Devices with Extremely Low Memory Footprint. 2020 IEEE Recent Advances in Intelligent Computational Systems,
RAICS 2020. https://doi.org/10.1109/RAICS51191.2020.9332480

[15] Mistree, K. B., Thakor, D., & Bhatt, B. (2023). An approach based on deep learning for Indian sign language translation. International Journal of Intelligent Computing and Cybernetics, 16(3). https://doi.org/10.1108/IJICC-08-2022-0227

[16] Mistree, K., Thako, D., & Bhatt, B. (2021). Towards Indian Sign Language Sentence Recognition using Insignvid: Indian Sign Language Video Dataset. Interna- tional Journal of Advanced Computer Science and Applications, 12(8). https://doi.org/10.14569/IJACSA.2021.0120881

[17] K, A., P, P., & Poonia, R. C. (2023). LiST: A Lightweight Framework for Continuous Indian Sign Language Translation. Information (Switzerland), 14(2). https://doi.org/10.3390/info14020079