International Research Journal of Education and Technology Peer Reviewed Journal

ISSN 2581-7795

Optimization of Sign Language Recognition using Deep Learning

Gokul S

[1]Studunt, Dept. of Artificial Intelligence and Machine Learning, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India

---***---

Abstract - *Sign language is an essential means of communication for individuals who are deaf or hard of hearing, enabling them to interact effectively with others. However, the lack of widespread understanding of sign language among the general population creates significant communication barriers. This results in difficulties in accessing essential services such as healthcare, education, and employment. While traditional sign language recognition systems have made progress in bridging this gap, many rely on specialized hardware, such as gloves or motion sensors, limiting their practicality and scalability in real-world settings. This research presents a deep learning-based approach to sign language recognition and speech-to-sign language conversion using computer vision techniques. The system integrates MediaPipe for real-time static gesture recognition and Long Short-Term Memory (LSTM) networks for dynamic sequence modeling, allowing for more accurate sign language interpretation. By leveraging publicly available datasets, including Indian Sign Language (ISL) and American Sign Language (ASL), the proposed system is designed to be scalable, adaptable, and capable of supporting multiple sign languages. By addressing these challenges, this research aims to create an inclusive, technology-driven solution that fosters better social integration and accessibility for individuals with hearing impairments. The proposed system has potential applications in various real-world scenarios, including educational institutions, workplaces, and public services, contributing to a more inclusive society where sign language users can communicate effortlessly with the broader population.*

Key Words: *Sign language recognition, deep learning, MediaPipe, LSTM, real-time communication, accessibility, Indian Sign Language (ISL), computer vision, gesture recognition.*

1.INTRODUCTION

1.1 Background of the Work

Sign language serves as the primary mode of communication for millions of people worldwide who are deaf or hard of hearing. However, despite its widespread use, sign language often remains an underrepresented form of communication in mainstream society. This disparity has resulted in communication barriers between deaf and non-deaf individuals, restricting access to essential services, education, and opportunities for social interaction. The need for seamless, real-time communication between both groups has become increasingly apparent, particularly in the context of technology and human-computer interaction. Traditional methods of sign language recognition have mostly relied on gesture-based input systems using manual sensors, gloves, or motion-capture systems. These approaches, although successful in controlled environments, are limited by practical constraints such as the need for specialized equipment, calibration requirements, and poor scalability. Moreover, they do not offer sufficient flexibility or portability to be useful in day-to-day life. The advent of deep learning and computer vision techniques, however, offers new opportunities to overcome these limitations, as these technologies can enable sign language recognition through real-time image or video processing, making it more practical for everyday use.

The focus of this research is to develop an efficient and accessible system that integrates both sign language recognition and speech-to-sign language conversion, thereby enabling real-time communication between deaf and non-deaf individuals. The project seeks to leverage deep learning models, particularly Mediapipe for gesture recognition and Long Short-Term Memory networks (LSTMs) for sequence modeling, to achieve an efficient sign language detection system.

1.2 Motivation and Scope of the Proposed Work

Communication barriers between deaf and non-deaf individuals remain a significant issue worldwide, hindering social inclusion and access to critical services. While sign language is a widely recognized language in the deaf community, its use is not universally understood by the broader population. As a result, people with hearing impairments often face challenges in daily interactions, including in healthcare, education, and employment.

While several technologies have been developed for sign language recognition, they often suffer from limitations such as poor accuracy, slow processing speeds, or dependence on specialized hardware. Current systems are typically designed for specific languages or require users to wear gloves or sensors, which may not be practical in real-world scenarios. Furthermore, existing speech-to-sign language conversion tools are not yet widely available, and many do not function in real-time or are limited to specific languages.

This project aims to address these challenges by developing a mobile application that performs both sign language recognition and speech-to-sign language conversion in real-time. The goal is to create an accessible, accurate, and efficient system that supports seamless communication between deaf and non-deaf individuals without requiring specialized hardware or setups.

2. METHODOLOGY

Sign language detection plays a crucial role in improving communication accessibility for individuals who rely on visual gestures. While various approaches have been developed, most conventional methods either require expensive hardware or suffer from poor real-time performance. Traditional sensor-based systems, such as gloves and motion sensors, provide precise movement tracking but lack portability and affordability. On the other hand, vision-based approaches using deep learning have shown significant promise but still face challenges in real-time processing, accuracy, and generalization across diverse users and environments. The need for an optimized, lightweight, and accurate sign language detection system has become increasingly evident as deep learning and computer vision continue to advance. To address these challenges, this research focuses on creating an efficient, real-time sign language detection system that integrates feature extraction and sequence modeling for enhanced performance.

2.1 Dataset Collection

The model is trained using publicly available datasets, including the Indian Sign Language dataset and the word-level American Sign Language dataset. These datasets contain labeled images and video sequences representing various gestures, enabling the training of a robust recognition model. Data augmentation techniques such as rotation, scaling, and background noise addition are applied to improve model generalization.

2.2 Feature Extraction

Mediapipe's Hand Tracking module is used to extract 21 hand landmarks per frame. These landmarks represent the x, y, and z coordinates of key hand joints, forming a structured dataset that captures hand movement dynamics. Compared to CNN-based feature extraction, Mediapipe significantly reduces computational costs while maintaining high accuracy.

2.4 Model Training

Training the LSTM model is a crucial step in ensuring accurate sign language recognition. Initially, a simple LSTM model was designed with moderate complexity, taking 30 frames of hand landmark data as input. The model consisted of multiple LSTM layers followed by dense layers for classification. However, during experimentation, it was observed that increasing the sequence length and modifying the architecture significantly improved accuracy. By increasing the number of LSTM layers and units, the model was able to capture temporal dependencies more effectively.

2.5 Deployment Phase

Once the model achieved satisfactory accuracy, it was prepared for deployment. The primary challenge in deployment was ensuring that the model ran efficiently in real-time while maintaining high accuracy. The trained LSTM model was saved using TensorFlow's SavedModel format, making it compatible with various deployment frameworks. To facilitate real-time interaction, the model was integrated into a lightweight web application using Streamlit. Additionally, OpenCV was used to process live video input, extracting frames that were fed into the trained model for prediction. The deployment phase also included performance optimization techniques such as quantization, reducing model size without significant loss in accuracy. The system was tested on different hardware configurations, ensuring smooth performance even on consumer-grade devices. By carefully handling latency and inference speed, the sign language recognition system was made practical for real-world use.

3. CONCLUSIONS

After training and deploying the word-level sign language detection model, multiple aspects of its performance and behavior were analyzed. Observations were recorded based on the effectiveness of feature extraction, sequence modeling, real-time inference, and system robustness under different conditions. The following sections present key observations categorized into various aspects of model performance and usability.

Fig 3.1 : Model Output

The research lays the foundation for future advancements in sign language recognition technology, emphasizing the need for continuous improvements in model generalization, user adaptability, and cross-linguistic support for different sign languages, including American Sign Language (ASL), Indian Sign Language (ISL), and British Sign Language (BSL). As machine learning models continue to evolve, integrating sign language detection into smart devices, augmented reality (AR) applications, and assistive communication tools will become increasingly feasible, enhancing the lives of sign language users worldwide

Suggestions for Future Work

While this research has successfully optimized sign language detection using deep learning and real-time processing techniques, several areas for improvement remain. Future advancements can focus on expanding dataset diversity, enhancing model robustness, and integrating multimodal learning approaches to further refine the system's capabilities.

1. Expanding the Dataset for Greater Generalization
2. Advancing Multimodal Learning for Context Awareness
3. Improving Real-Time Efficiency with Optimized Architectures
4. Extending the System to Sentence-Level Recognition
5. Enhancing the User Experience with Adaptive Learning

REFERENCES

[1] Y. I. Rokade, P. M. Jadav, "Indian Sign Language Recognition System," International Journal of Engineering and Technology (IJET), 2017. Available: https://www.researchgate.net/publication/318656956_Indian_Sign_Language_Recognition_System

[2] K. Shenoy, T. Dastane, V. Rao, D. Vyavaharkar, "Real-time Indian Sign Language (ISL) Recognition," arXiv preprint arXiv:2108.10970, 2021. Available: https://arxiv.org/pdf/2108.10970

[3] S. Patra, A. Maitra, M. Tiwari, K. Kumaran, S. Prabhu, S. Punyeshwarananda, S. Samanta, "Hierarchical Windowed Graph Attention Network and a Large Scale Dataset for Isolated Indian Sign Language Recognition," arXiv preprint arXiv:2407.14224, 2024. Available: https://arxiv.org/abs/2407.14224

[4] M. Rao, H. Kaur, S. K. Bedi, M. A. Lekhana, "Image-based Indian Sign Language Recognition: A Practical Review using Deep Neural Networks," arXiv preprint arXiv:2304.14710, 2023. Available: https://arxiv.org/abs/2304.14710

[5] P. Selvaraj, G. NC, P. Kumar, M. Khapra, "OpenHands: Making Sign Language Recognition Accessible with Pose-based Pretrained Models across Languages," arXiv preprint arXiv:2110.05877, 2021. Available: https://arxiv.org/abs/2110.05877