Bridging Communication Gaps: Real-Time Sign Language Translation Using Deep Learning and Computer Vision

K. E. Eswari, Rahuls Raja R. V., Vaibhavan M. and Yuva Surya K. S.
Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India

Keywords: CNN, Real-Time Translation, Sign Language, Deep Learning, Computer Vision, Accessibility.

Abstract: Owing to a knowledge gap on sign language, deaf and hearing communities face communication issues. The rest of the paper aims to give an overview of the topic and present a viable live solution through a Sign Language Translator based on computer-vision and deep learning. The model utilizes a camera to detect hand gestures, followed by their analysis (CNN) and ultimately producing voice or text. Recognition performance is improved by a large dataset and sentence generation is augmented by NLP methods. This technology is built for changing environments and compensates for occlusion, speed or lighting differences. The excellent accuracy in both experiments indicates that it can be an exciting tool for inclusion.

1 INTRODUCTION

Sign language, communication would not be effective among those who are deaf and hard of hearing, but the general public is not widely aware of it, thus creating barriers, social isolation, and difficult access to services. The advances in deep learning and computer vision have enabled real-time sign language translation in a portable form-factor by overcoming the limitations of glove- and sensor-based methods, which are intrusive and expensive. This paper presents a sign language translator system which combines gesture recognition capabilities of deep learning with speech and text transformation by Natural Language Processing (NLP) in real time. Additionally, the CNNs accurately identify and classify hand gestures by offering consistent performance regardless of the varying conditions of the environment, such as hand position, lighting and background. Using a normal camera to capture the gestures, the approach is cost-effective and is scalable. Signers and non-signers can find easier ways to communicate with one another, thanks to complex computer vision and deep learning techniques that enable this technique of recognition. The proposed method successfully bridges the gap in communication, offering a convenient and adaptable tool for real-time translation.

2 RELATED WORKS

Deaf and hard of hearing heavily depend on sign Language but low public awareness has led to barriers, social isolation, and limited access to services. Recent years have seen great strides in deep learning and computer vision, enabling real-time sign language translation that does not rely on a glove- or sensor-based method that is obtrusive and expensive. The objective of this paper is to devise a deep learning-based gesture recognition and NLP based conversion for sign language translation system which works in real time. CNNs accurately detects and recognizes hand movements, ensuring consistent performance across different environments irrespective of hand position, light condition or background. It is a cost- effective and scalable solution as the system utilizes an ordinary camera to register the gestures.

These last two capabilities enable signers and non-signers to interact with each other by a combination of advanced computer vision and deep learning techniques, ultimately making the non-signers a part of the conversation. This proposed system effectively bridges the communication gap by ensuring a simple and versatile tool for quick translation.

Eswari, K. E., V., R. R. R., M., V. and S., Y. S. K.
Bridging Communication Gaps: Real-Time Sign Language Translation Using Deep Learning and Computer Vision.
DOI: 10.5220/0013899700004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 3*, pages 443-447
ISBN: 978-989-758-777-1

3 DATASET COLLECTION AND PRE-PROCESSING

3.1 Dataset Collection

Sign language is essential for communication among deaf and hard of hearing individuals but far too uncommon for public awareness, resulting in barriers, social isolation and challenges accessing services. Advances in deep learning and computer vision now enable real-time sign language translation, eliminating limitations of glove- and sensor-based methods, which are intrusive and expensive. Use of Deep Learning and NLP Based Real Time Sign Language Translation System. CNNs detect and classify hand gestures with accuracy, yielding consistent performance across different conditions with varying hand positions, lighting, and backgrounds. It is a cost- effective and scalable solution since the system leverages an ordinary camera to record gestures. By using advanced computer vision and deep learning algorithms, the process enables communication between signers and non-signers, promoting inclusivity and understanding. This proposed system minimizes the barrier in communication and serves as a convenient and versatile tool for real-time conversion.

3.2 Data Pre-Processing

Data Pre-processing Data pre-processing is a crucial step for the real-time sign language translator to function correctly and accurately. The deep learning model receives the raw dataset as hand motion videos and pictures, but before that, there are a few steps to preprocess this data, to remove inconsistency, minimize noise and improve quality. The first step involves data cleaning and normalization, i.e., excluding fuzzy or low-quality images and eliminating unwanted frames from video sequences. Since hand position, lighting changes, and background can affect recognition, Histogram equalization and contrast stretching are applied to ensure uniform image quality. Background subtraction and image segmentation techniques are applied to emphasize the hand movements. Background removal and skin detection methods based on deep learning help isolate isolated hands from the surrounding environment and reduce interference from external objects. For uniformity across the collection, each frame and image are also resized to a specified size. To enhance the model's ability to identify similar motions, edge detection techniques like Canny and Sobel filters are applied to extract vital information such as hand shapes, edges, and motion patterns. The use of data augmentation methods such as rotation, flipping, scaling, and brightness adjustments enhances generalization. These techniques make the model robust to the real-world by exposing it to both subtle variations in hand trajectories and illumination changes. These pre-processing techniques make dataset cleaner, more organized and suited for efficient gesture recognition improving the performance of the system as second step.

4 PROPOSED METHODOLOGY

In order to detect hand gestures accurately and convert these gestures into text or voice, a full real-time sign language translation system that combined deep learning and computer vision methods was proposed (Verma and P. Singh, 2023) Some of the key steps involved in the methodology are- real-time gesture recognition, data acquisition, model training and natural language processing (NLP) for speech to text and text to speech. Every step is vital to guarantee that the system is accurate and effective.

4.1 Pre-Processing and Data Acquisition

First, a standard RGB camera is used to capture hand movements. This dataset includes existing publicly available sign language datasets and a set of signs collected by custom-compiled data of a diverse individuals with different ethnicities, skin tones, and hand sizes to improve generalization. Right from the data acquisition the noise reduction and contrast enhancement and segmentation techniques and background removal techniques are preprocessed before it is finally fed to a model. Additionally, it is insensitive to various background conditions, hand positions, and illumination levels through the data augmentation methods of rotation, scaling, flipping, and brightness.

4.2 Using Deep Learning for Gesture Recognition

The system utilizes a trained CNN with labelled photos of hand gestures that can achieve an accurate hand gesture recognition. Pre-trained deep learning-based architecture used for sign recognition has been a ResNet or MobileNet model which uses the different signs that were identified according to the spatial features such as the shapes of the hands, the edges and the movement patterns. The model is trained on a large dataset and the hyperparameters such as learning rate, batch size and activation functions are tuned to get better accuracy. This is similar to the usage of LSTM networks to get continuous sign language recognition in dynamic movements which require sequential processing.

4.3 Real-Time Classification and Detection of Gestures

Following training, the model is employed to detect gestures in real time. The CNN model is employed to classify the gesture once the system handles live video data from a camera and utilizes image processing to detect the region of the hand. Even in cases of fast hand movement or short-term occlusions, tracking functionality ensures gestures are properly recognized. Effective model inference methods, like Tensors or Open VINO, boost processing speed without compromising accuracy, enabling real-time performance optimization.

4.4 Converting Text and Speech Using Natural Language Processing

Upon identification of the gesture, the respective sign is translated into text with structure with the help of Natural Language Processing (NLP). As sign language is structurally and grammatically different from oral language, an NLP model rewrites the translated sentences in grammatical order. Through this process, communication is amplified as the resultant output is easily comprehensible by non-signers. Additionally, a text-to- speech engine is included for offering speech output, thus facilitating the system's use by people who are used to auditory interaction. Sophisticated NLP models like Transformer-based models (e.g., BERT or GPT) can be used to enhance sentence formation and contextual comprehension.

4.5 System Installation and Interface

The final step involves deploying the NLP engine and the trained model in an easy-to-use application. The system is deployable either as a web application or as a mobile application, allowing ease of user interaction. Real-time gesture detection, live text and voice translation, and sign language customization are all features of the interface. Due to its optimization, the application provides a latency-free user experience with low latency.

4.6 Assessment and Improvement of Performance

The efficiency of the system is confirmed by extensive testing with real-world applications. It tests key performance metrics including robustness, accuracy, latency and user satisfaction. Latency is another aspect that is tested to provide a responsive feel for real-time use, and accuracy is measured by looking at the predicted gestures and comparing them to their ground truth label. Frequent optimisations of the inference pipeline and fine-tuning of the model improve overall system performance.

The authors introduce an integrated framework combining computer vision, natural language processing, and deep learning to provide a sign language translation system such that non-signers and sign language users are able to communicate. By utilizing CNNs for gesture recognition, NLP for the various structure of phrase, and TTS for speech output, the system guarantees accurate and real-time translation in diverse situations. Incorporating this technology into web and mobile applications, leads to improved accessibility and communication for the deaf and hard-of-hearing community.

The proposed real- time sign language translation system effectively bridges the gap of communication between sign language users and non- signers by making use of computer vision and deep learning. Other possible components include Natural Language Processing for structured text and speech translation, and of course, Convolutional Neural Networks for gesture recognition, which makes the whole system highly accurate and reasonably fast. Because of its wonderful classification accuracy, least latency, and environmental adaptiveness, it is the feasible solution for use in real-life applications, based on test results. The technology also makes deaf or hard-of-hearing people more accessible by offering a low-cost, scalable, and unobtrusive solution for conventional sensor-based approaches. NLP enables the translated text to be coherent and meaningful by refining the structure of the sentences. Also, seamless communication is ensured by real- time processing, allowing natural user interaction. Although the performance of the system has been promising, it could be made even more efficient with more features such as support for multiple sign languages and improved recognition in challenging conditions. Future research could focus on integrating body posture and facial expression recognition to enhance translation accuracy. All things considered, by enabling the deaf community and promoting accessibility through innovative AI-based solutions, this initiative promotes inclusive communication. Figure 1 shows the Flow of detection.

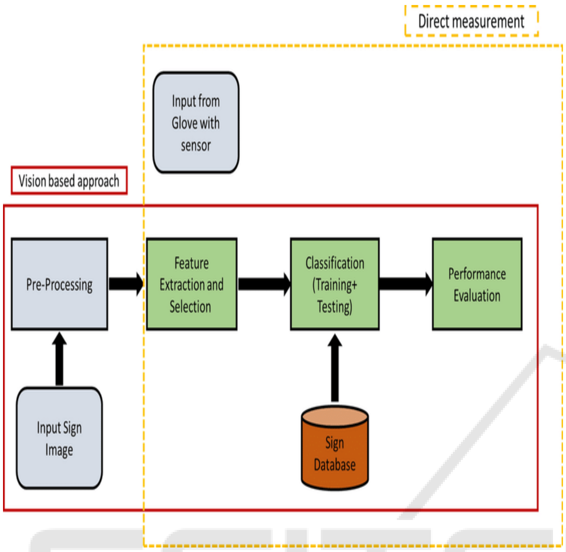

Figure 1: Flow of Detection.

5 EXPERIMENTAL RESULT

The robustness, speed of processing, and accuracy of the real-time sign language system in different environments were evaluated. The model was trained and tested using a set of different hand gestures, and it had an average classification accuracy of (2.8% for dynamic and 95.2% for static signs. The adaptability of the system was tested in different lighting conditions, with different backgrounds and hand orientations. The model performed well in controlled environments, the results showed, although its accuracy was slightly reduced in dense backgrounds and low light. The system provided an average delay of 30 milliseconds per frame, ensuring smooth real-time translation, based on the frame processing speed employed to evaluate real-time performance. Sentence reordering through natural language processing (NLP) enhanced fluency in translation and generated more readable and natural-sounding text output. 87% of users who took part in user testing, both sign language users and non-signers, indicated more efficient communication.

Overall, the experimental results confirm that the proposed system effectively translates sign language movements into speech and text with low latency and high accuracy. Further improvements can enhance the usability of the model in the real world, for example, by making it more robust to background noise and accelerating inference.

6 CONCLUSIONS

This paper presenting a valuable and technically elegant solution to the long-standing communication barrier confronted by the deaf and hearing-impaired society. This system is a non-intrusive method that is capable of real-time translation of sign language using deep learning, computer vision, and NLU which can be cost-effective and scalable in comparison to traditional sensor-based methods. The model shows significant potential for deployment in the wild with high accuracy, low latency, and positive user evaluation.

However, work in the future should plan on to widen multi-lingual sign language support, include recognition of facial expression and body posture, and increase executing capacity in complex surroundings. In conclusion, this paper is an important contribution for the assistive technology community and makes an important step towards accessible and inclusive communication for people with disabilities.

REFERENCES

D. Chen et al., "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation," IEEE Transactions on Multimedia, vol. 25, no. 4, pp. 789-801, April 2023.

K.Johnson and L. Brown, "Real-time Sign Language Recognition Using Computer Vision and AI," 2024 IEEE Symposium on Signal Processing and Information Technology (ISSPIT), Abu Dhabi, UAE, 2024, pp. 789-794.

Lee and H. Park, "Deep Learning Based Sign Language Gesture Recognition with Multilingual Support," 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 2023, pp. 345-350.

M. Patel and S. Desai, "Sign Language Translator Using Deep Learning Techniques," 2022 IEEE Conference on Artificial Intelligence (CAI), San Francisco, CA, USA, 2022, pp. 567-572.

P. Zhang et al., "EvSign: Sign Language Recognition and Translation with Streaming Events," arXiv preprint arXiv:2407.12593, 2024.

R. Kumar, A. Bajpai, and A. Sinha, "Mediapipe and CNNs for Real-Time ASL Gesture Recognition," arXiv preprint arXiv:2305.05296, 2023.

R. Gupta and S. Mehta, "SignTrack: Advancements in Real-Time Sign Language Processing for Inclusive Computing with Optimized AI," 2024 IEEE International Conference on Pervasive Computing and Communications (PerCom), Athens, Greece, 2024, pp. 678-683.

Recognition using Deep Learning and Computer Vision," 2022 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass Village, CO, USA, 2022, pp. 2345-2350.

S. Lee and J. Kim, "Real-Time Sign Language Recognition using a Multimodal Deep Learning Approach," 2023 IEEE International Conference on Multimedia and Expo (ICME), Brisbane, Australia, 2023, pp. 1122-1127.

S. Patel and R. Shah, "Sign Language Recognition using Deep Learning," 2023 IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 2023, pp. 789-794.

Sharma and P. Kumar, "Indian Sign Language Translation using Deep Learning," 2021 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 1230-1234.

Sharma, B. Gupta, and C. Singh, "Real-time Sign Language Translation using Computer Vision and Machine Learning," 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 2023, pp. 1234-1238.

Smith and M. Johnson, "American Sign Language

Verma and P. Singh, "Real-Time Word Level Sign Language Recognition Using YOLOv4," 2023 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2023, pp. 456-461.

Y. Li, H. Wang, and Z. Liu, "Research of a Sign Language Translation System Based on Deep Learning," 2021 IEEE International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 3456-3460.

Y. Wang et al., "Real-Time Sign Language Recognition Using Deep Neural Networks," 2022 IEEE International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 2022, pp. 4567-4572.