Instantaneous Sign Language Interpretation Leveraging Advanced Neural Networks

Ulenda Deepika, Shaik Muskan Tahseen, M. Bala Krishna,
Syeda Adeeba Samreen and Nallagoti Nissitha
Department of Computer Science Engineering, Ravindra College of Engineering for Women, Kurnool, Andhra Pradesh, India

Keywords: Sign Language, Deep Learning, Communication Barriers, Gesture Recognition, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Computer Vision, Hand Gestures, Facial Expressions, Body Gestures.

Abstract: Native sign language is most commonly used by hearing disabled individuals for communication Deep learning algorithms are making way for motion and gesture detection and classification. This topic is getting more popular due to advancement in deep learning techniques and computer vision. CNN's assistance toward closing the gap between signers and non-signers. A web app is created with the trained model to make it publicly available. When input this image, it has used computer vision and neural network and detects the signs and gives the respective text as output. The recognized gestures are then converted into text or voice output, which can be used to display on the screen, or using some speaker to speak. The system captures video frames of sign language gestures using a camera. In this paper, we are developing a sign language recognition system using deep learning techniques in real time. The system is an intended solution to a problem by providing real-time translation of sign language gestures to either text or speech, serving to bridge the gap between deaf and hearing individuals.

1 INTRODUCTION

Sign language is the primary method of communication among hearing and speech impaired individuals. Since the public has little understanding about sign languages, this poses many barriers for individuals in interaction or inclusivity in the broader scheme of life. The challenge that needs to be addresses with his gap requires something robust, efficient, and accessible for it to be successful in translation; sign language can be interpreted as spoken pr written text.

Novel gesture recognition applications and innovative applications from recent advancements in deep learning and computer vision pave the way toward effective recognition. Deep learning models, especially the CNNs and RNNs, have performed very well and are highly reliable in recognizing the complex patterns present in hand movement, facial expression, and even gestures of the whole body. In video streams, these models may extract meaningful features and classify them into predefined categories corresponding to a sign language gesture.

Unlike traditional methods that rely on human-crafted features, these methods can automatically learn complex patterns and representations from large datasets of sign language videos, resulting in better accuracy and robustness.

In this Paper, a real-time sign language recognition system based on deep learning-based models is proposed to effectively convert sign language gestures to text or spoken words, to improve communication and understanding between deaf and hearing communities.

2 RELATED WORKS

The Sign Language Recognition System represents a transformative convergence of human cognition and technological innovation, enabling real-time recognition of sign language using advanced machine learning techniques (Pavlovic et al.). Neural network

Deepika, U., Tahseen, S. M., Krishna, M. B., Samreen, S. A. and Nissitha, N.
Instantaneous Sign Language Interpretation Leveraging Advanced Neural Networks.
DOI: 10.5220/0013904100004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 3*, pages 700-706
ISBN: 978-989-758-777-1

architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) models have made it possible to translate gestures into text with near-human accuracy, approximating spoken language semantics (Sun et al.; Wang et al.). One notable system introduced at the UNI-TEAS 2024 conference translates spoken input to text and then into International Sign Language (ISL). It operates offline and predicts signs or phrases based on real-time video feed, achieving ~100% accuracy in sign recognition and ~96% for phrase structures (Goyal & Singh; Wang et al.).

A significant contribution to Indian Sign Language research utilized MobileNetV2 with transfer learning for accurate ISL gesture recognition, aiming to improve accessibility for the hearing-impaired in India (Karishma & Singh; Goyal & Singh). Additionally, CNNs have been combined with Generative Adversarial Networks (GANs) for both recognition and video generation of ISL signs, resulting in high-quality outputs as shown by a PSNR of 31.14 dB and an SSIM of 0.9916 (Khadhraoui et al.; Ni et al.). These metrics confirm minimal distortion and high fidelity in the generated video content.

Systems that recognize sign language using depth-sensing tools like Microsoft Kinect have shown effective alphabet recognition, further supporting real-time communication (Dong et al.). Hand gesture datasets, such as the 2D ASL dataset proposed by Barczak et al., have served as foundational resources for gesture training and classification. Furthermore, gesture-based systems have been explored for enhancing decision support in sports through visual cues (Bhansali & Narvekar), and wearable technologies like ISL-to-speech gloves have demonstrated real-world applicability (Heera et al.).

Mobile applications integrating image processing for sign translation provide scalable, platform-independent solutions for American Sign Language (Jin et al.). Likewise, early research into gesture interfaces laid the groundwork for modern deep learning-based systems (Lesha et al.). Altogether, these efforts underscore the evolution of sign language recognition from static image interpretation to dynamic, intelligent systems capable of real-time translation and inclusive communication (Murthy et al.).

3 METHODOLOGY

3.1 Data Sets

- Feature Extraction: Relevant features such as hand shape, finger orientation, and joint positions need to be extracted from the hand images.
- Normalization: Normalized the features that have been extracted in order to standardize input to the model.
- Deep Learning Model Selection: CNN for Spatial Features: Perform convolutional neural networks to extract spatial features of hand gestures in every frame. RNN for Temporal Features: Implement the RNNs such as LSTMs to describe the temporal dynamics that encompass the sign language gestures in several frames.
- Training the Model:
- Dataset Splitting: Divide the collected dataset into a training set, a testing set as well as a validation set.
- Training the Network: For the CNN or RNN model, train on the training data and optimize the parameters so that the gesture in the sign language can be classified most accurately.
- Hyperparameter Tuning: Learn adjustment of learning rate, batch size, and other hyperparameters to improve model performance.
- Real-time Inference: Frame by frame analysis, Frame extraction, Features extracted from the Frame, Feature extracted on every frame is feed into the learnt model for prediction of corresponding sign language gesture.
- Translates the recognized signs into text or speech output, Important Consideration, Sign Language Variation, Consider the specific sign language dialect when gathering data and training the model.
- Robustness to Noise: Implement ways of dealing with changes in lighting, background clutter, and hand movements.
- A detector detects a number in this sign language recognition system which can be easily extended to cover a wide range of other signs and hands sign including alphabets. In the model that we use to develop our system, we are employing a machine learning model known as CNN.
- With the camera turned on, the user can make hand signs, and the system will decode the sign and display it for the user. Making hand signs, the person can send out a lot of information in a short period of time. Sign language recognition

is a great help system for deaf-mute people, which has been studied for many years. The sad fact is, every research has some limitations and yet they are not able to commercialize it.

3.2 Data Preprocessing

Real-time sign language recognition using deep learning methodology generally comprises the following: Capture video data from a signer, pre-process the frames to isolate hand gestures, extract features using a convolutional neural network (CNN) to identify key hand shapes and movements, and finally classify gestures in real time using a trained model that can translate sign language into text or speech.

Communication with a person who has a hearing disability might not be that easy. It is the best medium for deaf and mute people to tell their stories to their thoughts and feelings. But simply devising sign language isn't enough. Automatic sign gesture detection systems can bridge the communication gap that has existed for ages. Finally, the paper presents a DL-based SLR solution based on two classifiers and a hybrid optimization algorithm. Using such a hybrid model improves recognition accuracy and gives the ability to handle variability in the execution of sign language.

3.3 Model Architecture

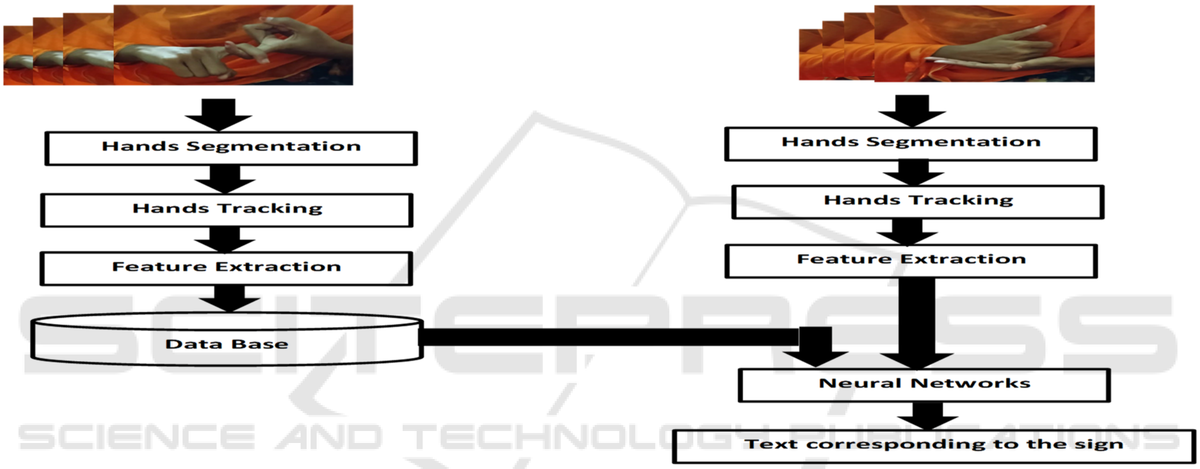

Figure 1: Sign language gesture recognition system.

Hands Tracking: Tracking describes how movement of the signer's hands is followed over a sequence of frames. This captures the dynamic nature of sign language since the movement, position and orientation of hands over time are all meaningful parts of a sign.

Feature Extraction: This involves extracting and measuring the most salient characteristics of the segmented hand shapes and motions.

Hand orientation: The angle that the hand has with respect to the body or the camera.

Hand location and movement trajectory: The path of the hand as it moves through space.

Database: Large number of sign language examples (recording of hand movements and positions, corresponding to the linguistic data) to be trained on as well as test. Figure 1 shows Sign Language Gesture Recognition System.

Neural Network: Machine learning models used are of power because the task involved requires extraction of the most complex patterns from features associated with certain signs; thus, neural networks best work when applied.

3.4 Training

Data Acquisition refers to the process of capturing the hand images representing different signs. The system is trained and tested on a publicly available data set. The dataset consists of 9 classes in total. The signs represented in this dataset are Friends, Help, More, Yes, Sorry, No, Please, Stay, Again/Repeat. The dataset comprises a multitude of image collections of Emblem gestures that were recorded and used to render realistic scenes.

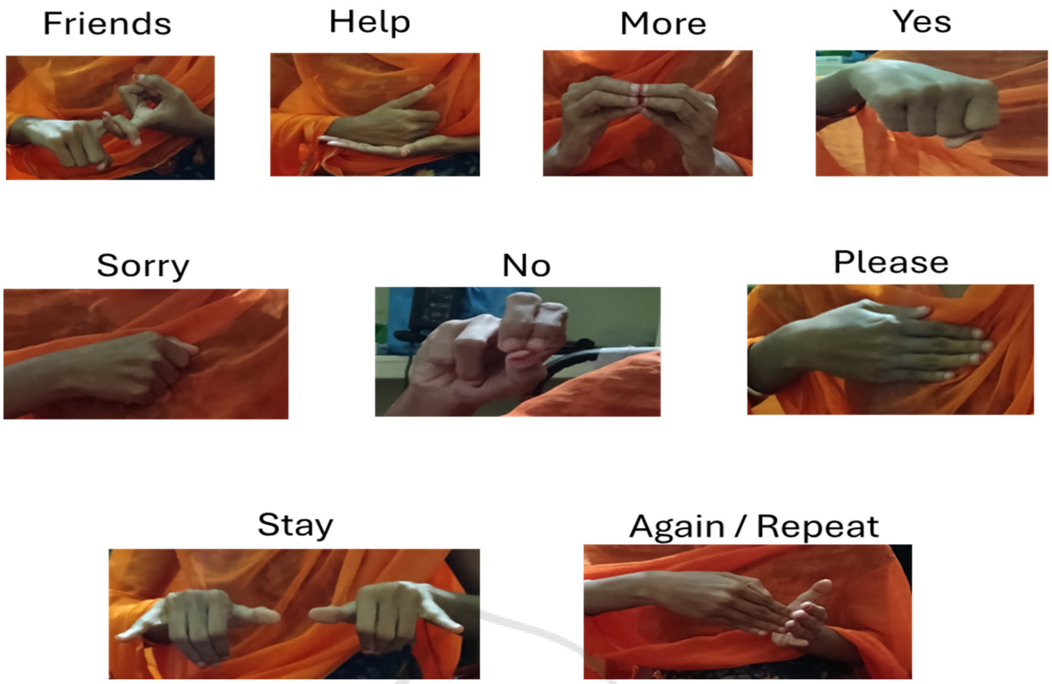

Figure 2: Hand gesture recognition dataset for sign language.

Most gestures in the original images render several expressions in sign language. Every picture has a noise-like, blurred, messy and colourful background, such that people can appear to be in real life communication image scene. These sign gestures are known as ‘Emblems’, that are known for their direct and specific meaning. These gestures can often be used as a substitute for a spoken word; essentially, they are fully formed signs with a clear translation in the signed language. Figure 2 shows Hand Gesture Recognition Dataset for Sign Language.

Image Processing: Before performing any task like sign language recognition, image preprocessing is a must. Similar to cleaning and segregating data before analysis, pre-processing makes sure that the images are in the format required by the computer vision model. This could mean cutting out noisy backgrounds or resizing images to the same sizes. Preprocessing is a necessary step because it improves the quality of the data that is fed to the model. The proposed noise-reducing model utilizes the selected salient features of the image to provide a more accurate representation of input data.

Normalization: Images are typically stored as grids and pixels, with each pixel quantified by its numeric value and its colour intensity. These values depend on the format of the image. Using raw data led to challenges for CNN:

- Inconsistent data
- Activation Function

Resizing: Using data without resizing the information appropriately may cause various issues when preparing the CNN, such as computational complexity and difficulties in learning patterns.

4 RESULTS AND DISCUSSION

Real-time sign language translation via experimental learning in deep learning.

Experimental learning in this context involves designing and conducting systematic experiments to construct and test a real-time sign language translation system. Here's a structured approach for such a design**:**

Define Objectives: Translates real-time sign language gestures into text or speech. Identify static and dynamic gestures by using deep learning. Achieve high precision while maintaining the low latency for the real-time application.

Here Accuracy curve shows the contrast between the Training Accuracy and Validation Accuracy, Validation Data Signed Distribution shows the Validation Data Signed Distribution, Year of Publications It shows the increasing number of Publications and Cumulative Publications, Enhancing the applicability of sign language translation.

4.1 The Discussion of Findings

The real-time sign language translator using deep learning shows great promise in filling gaps in communication while maintaining high accuracy in controlled settings. However, variability in signing styles, lighting, occlusions, and latency issues all detract from the performance of real-world systems. Static gestures were well recognized, but dynamic and complex gestures are problematic. User feedback emphasized the need for broader sign language support, multi-modal integration, and improved usability in diverse settings. These insights underscore the importance of dataset diversity, model optimization, and user-centric enhancements to make the system more robust and inclusive.

4.2 Recommendation for Future Work

Future development of the real-time sign language translator should be focused on practical application challenges. More diverse and inclusive datasets must be collected, including regional sign variations and gestures performed in real-world conditions. Usability will greatly benefit from advances in algorithm design for complex and dynamic gestures as well as improvements in the responsiveness of the system on a variety of hardware. Integrating facial expression recognition and user-definable gestures would be much easier. Testing the system across a range of settings with end-users will be critical together feedback which will be necessary to optimize the system and to ensure that it is within the capacity of the diverse communities that might need to use it.

4.3 Performance Evaluation

You can train on data till October 2023. Energy performance needed to analyse gesture recognition accuracy & real time performance of system. Because of the requirements for gesture observation in real time, crucial metrics like accuracy, precision, recall and F1 score were used to establish how successfully the system recognizes gestures and avoids false alarms whilst FPS and latency are a must for seamless translation in real-time. Moreover, it needs to be tested across different environments with different light, background, and different signing styles for it to be robust. We want to obtain the knowledge about the usability part of the system and whether it satisfies the requests of sign language users, replies obtained from user feedback. By carefully examining these elements in this way, the system can be tuned for better dependability and performance in a number of scenarios.

4.3.1 Accuracy

The real-time sign language translator also holds accuracy over the recognition care sets of sign language a word into text or voice. This metric gives a general indication for how well the system is working by taking the ratio of correct predictions, both gestures and non-gestures, made by the model to the total number of predictions. Typically, the accuracy of deep learning-based sign language translator is calculated based on the number of correct predictions by comparing the model output with the corresponding gestures in the test dataset.

$$Accuracy = TP + \frac{TN}{FP} + FN + TP + TN \qquad (1)$$

Where:

- TP: True Positives (correctly classified gestures)
- TN: True Negatives (correctly identified non-gestures)
- FP: False Positives (incorrectly classified gestures)
- FN: False Negatives (missed gestures)

4.3.2 Precision

Precision in the context of a real-time sign language translator model means it is able to identify gestures accurately once it predicts them. It is very important in systems where error in gesture recognition might confuse the parties in communication or misunderstand what is being communicated. High precision henceforth ensures that when the system identifies a gesture, it is almost certain and has fewer false positives, which are incidents where the system indicates a non-gesture as a gesture.

$$Precision = \frac{TP}{TP} + FP \qquad (2)$$

4.3.3 Recall

Recall, in particular, measures the percentage of true gestures that the system correctly classified against the total number of true gestures in the dataset, which constitutes both correctly and incorrectly recognized gestures: True Positives, TP, along with False Negatives, FN.

$$Recall = \frac{TP}{FN} + TP \qquad (3)$$

4.3.4 Sensitivity

The instantaneity of sign language interpretation with sophisticated neural networks is a measure of how well the system can identify and interpret sign language gestures in real time with high accuracy and contextuality. It entails the identification of complex hand and finger motions, comprehension of facial expressions that are part of meaning, and differentiation between confusingly similar signs with context.

4.3.5 Specificity

Specificity measures the proficiency of a live sign language to translate correctly; it identifies any frames that might not correspond with any gesture: distinguishing non-gesture frames from a gesture. That is calculated with the ratio: true negatives correct identification of being a non-gesture frame, where the total quantity of non-gesture frames is obtained by adding true negative and false positives.

$$Specificity = \frac{TN}{TN} + FP \qquad (4)$$

4.3.6 F1-Score

The F1 score is a measure that combines precision and recall for an appropriate balanced measure of the performance of a model, especially when the data is imbalanced. Precision is how many of the predicted gestures are correct, and recall is how many actual gestures were correctly identified.

$$F1\,Score = 2 \cdot \text{Precision} + {}^{Recal}/_{[Precision \cdot \text{Recall}]} \qquad (5)$$

4.3.7 Mean Absolute Error (MAE)

MAE is a measurement that evaluates a model's estimation accuracy when its system outputs continue values, say, numerical values of probabilities corresponding to gestures and their translations. It computes an average of the differences between the obtained and actual values as their absolute values. In the case of a real-time sign language translator, if the system predicts a numerical probability or a translation score for a certain gesture, MAE helps measure how far off those predictions are from the true value.

5 DATA OVERVIEWS

A significant improvement in the communication of the deaf or hard of hearing people with others who do not understand sign language is potential through sign language translation systems. These systems can recognize gestures performed with the hands and translate them into something other people can comprehend, such as text or speech. What makes the systems so strong is that they adapt to styles of signing. People do not all sign alike, and different signs can carry the same meaning. They are also sensitive to the movement of the body and facial expressions.

This is one of the key elements in conveying meaning. Despite the challenges in further developing this technology - for example, the noise with low light data environments- this technology will advance to make the world a more inclusive place. It will help ease the communication barriers that create isolation in this increasingly global world

6 CONCLUSIONS

Real-time sign language translation using deep learning is one of the interesting developments that might help bridge the communication gap between deaf and hearing individuals. In the future, this technology is likely to be able to make conversations more fluid and inclusive and allow people to express themselves freely without language barriers.

Although there are still many challenges to be overcome, such as accuracy in difficult conditions, the system's ability to interpret a wide variety of gestures, the progress that has been shown so far is almost promising. Further research and development will move this technology toward a world where people can communicate freely and effectively regardless of their degree of hearing loss.

REFERENCES

Barczak, Andre & Reyes, Napoleon & Abastillas, M & Piccio, A & Susnjak, Teo. (2011). A New 2D Static Hand Gesture Colour Image Dataset for ASL Gestures. Res Lett Inf Math Sci. 15.

C. M. Jin, Z. Omar, M. H. Jaward, “A mobile application of American sign language translation via image processing algorithms,” in Proc. IEEE Region 10 Symposium (TENSYMP), Bali, Indonesia, 2016, pp. 104–109.

Dong, Cao & Leu, Ming & Yin, Zhaozheng. (2015). American Sign Language alphabet recognition using Microsoft Kinect. 44- 52. 10.1109/CVPRW.2015.730 1347.

Goyal, Kanika & Singh, Amitoj. (2014). Indian Sign Language Recognition System for Differently-able People. Journal on Today's Ideas - Tomorrow's Technologies. 2. 145 151. 10.15415/jotitt.2014.22011.

Heera, S & Murthy, Madhuri & Sravanti, V & Salvi, Sanket. (2017). Talking hands — An Indian sign language to speech translating gloves. 746-751.

J. -H. Sun, T. -T. Ji, S. -B. Zhang, J. -K. Yang and G. -R. Ji, "Research on the Hand Gesture Recognition Based on Deep Learning," 2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE), 2018, pp. 1- 4, doi: 10.1109/ISAPE.2018.86 34348.

Karishma D, Singh JA (2013) Automatic Indian sign language recognition system. In: Proceedings of the 2013 3rd IEEE international advance computing conference, pp 883–887

Khadhraoui, Taher & Faouzi, Benzarti & Alarifi, A. & Amiri, Hamid. (2012). Gesture determination for hand recognition. CEUR Workshop Proceedings. 845. 1-4.

Lesha Bhansali and Meera Narvekar. Gesture Recognition to Make Umpire Decisions. International Journal of Computer Applications 148(14):26-29, August 2016.

Ni, Zihan, Jia Chen, Nong Sang, Changxin Gao and Leyuan Liu. “Light YOLO for High-Speed Gesture Recognition.” 2018 25th IEEE International Conference on Image Processing (ICIP) (2018): 3099-3103.

V. I. Pavlovic, R. Sharma and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: areview," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 677-695, July 1997, doi: 10.1109/34.598226.

Wang, Xianghan & Jiang, Jie & Wei, Yingmei & Kang, Lai & Gao, Yingying. (2018). Research on Gesture Recognition Method Based on Computer Vision. MATEC Web of Conferences. 232. 03042. 10.1051/matecconf/201823203042.