© JUN 2023 | IRE Journals | Volume 6 Issue 12 | ISSN: 2456-8880

Signassist: Sign Language Interpreter Using Deep Learning

KARAN KHARBANDA[1], UTSAV SACHDEVA[2], PROF. DR. ANU RATHEE[3]
[1, 2] Student, Maharaja Agrasen Institute of Technology, Delhi, INDIA
[3] Professor, Maharaja Agrasen Institute of Technology, Delhi, INDIA

Abstract- Effective communication is essential for individuals to express their ideas and emotions. However, persons with speech or hearing disabilities often face significant communication barriers. To address this issue, deep learning models, specifically LSTM and GRU, are proposed to recognize and translate signs from isolated American Sign Language (ASL) video frames. In this research, transfer learning and data augmentation techniques are utilized to develop a deep learning model for the ASL dataset. The proposed models achieve up to 95% accuracy in recognizing signs from ASL datasets. This research aims to develop a more natural and efficient way of communication for persons with hearing impairments and promote collaboration with people not trained in sign language. Overall, this study demonstrates the potential of deep learning models to reduce communication barriers and promote inclusivity.

Indexed Terms- American Sign Language, SignAssist, Deep Learning, Sign Recognition, Gesture Recognition

I. INTRODUCTION

Sign language is a critical mode of communication for individuals who are deaf or hard of hearing. However, it can be challenging for non-signers to understand sign language, making trained sign language interpreters necessary for various medical, legal, and educational sessions. Researchers have developed video remote human interpreting and automated American Sign Language (ASL) recognition studies with the growing demand for translation services. However, these studies rely on traditional shallow neural network approaches that require manual feature selection or are only tested on small datasets.

Deep learning (DL) methods have shown significant improvements in machine learning applications, particularly in image recognition and computer vision issues. Thus, this paper proposes an automated sign language translation system using DL methods that can be used to facilitate communication between speech-impaired individuals and non-signers without the need for an interpreter. The proposed system utilizes Image Segmentation and Object Tracking techniques to isolate and analyze images and detect and track moving objects.

By automating sign language translation, this system can significantly improve communication for speech-impaired individuals and reduce their dependence on sign language interpreters. This would help speech-impaired individuals accurately convey their symptoms to a doctor, ask questions during educational sessions, and communicate effectively in various settings. Thus, the proposed system has the potential to make a significant impact on the lives of speech-impaired individuals globally. Our system makes use of an external web camera or a built-in camera in the computer for detecting and tracing movements and landmarks of the hands.

The proposed system provides a promising solution to the communication barriers faced by speech-impaired individuals globally. It utilizes deep learning methods to automate sign language translation, reducing dependence on sign language interpreters and improving communication with non-signers in various settings. Further research in this field could lead to significant advancements in communication technology and help to create a more inclusive society for those with speech and hearing impairments.

A. Problem Description and Overview

Human interaction is an important aspect for humans to communicate and is primarily done through speech, but those with speech impairments must rely on tactile-kinaesthetic communication methods. Sign language, also known as visual language, is a valuable tool for those with speech and hearing impairments, with basic parameters such as hand shape, orientation, movement, location, and components like mouth shape and eyebrow movements.

The system proposes an automated sign language interpretation system that utilizes deep learning methods to facilitate communication between speech-impaired individuals and machines without the need for a facilitator. By utilizing Image Segmentation and Object Tracking techniques to isolate and analyze images and detect and track moving objects, this system can significantly improve communication for speech-impaired individuals.

B. Objective

The objective of SignAssist is to make a significant impact on the lives of speech-impaired individuals globally, allowing them to accurately convey their symptoms to a doctor, ask questions during educational sessions, and communicate effectively in various settings. It addresses the issue of communication barriers faced by those with speech and hearing impairments and presents a solution that captures video of hand gestures, processes it, and passes it to a model that predicts words and generates a meaningful sentence in the selected language to remove these barriers.

II. LITERATURE REVIEW

Related studies have been undertaken and worked upon in this domain. The proposed system is built upon those developments and they act as the backbone of our mechanism.

American Sign Language is a visual-spatial language developed in America. American Sign Language is a natural language with its own phonology, morphology and grammar. It uses the arms, hands, face, and body/head to construct semantic information that expresses words and emotions [4]. Nandy et al. [5] proposed a method for detecting and recognizing hand gestures from grayscale images in English. In their method [5], video segments containing motions are converted into grayscale frames from which features are extracted using a direction histogram. Finally, clustering is used to classify signs into one of the predefined categories based on their properties. The authors achieved 100% numerical recognition in their study and concluded that the 36-thousand histogram method is more accurate than the 18-thousand histogram method. Mekala et al. [6] proposed a neural network to recognize and track signatures and generate real-time text from video streams. System architecture, framing, image preprocessing, feature extraction based on hand points and meshes, etc. It has many levels such as These cell attributes are represented using points of interest (POIs) in the cell [6]. Using this method, 55 different features were extracted, which the authors proposed were used as inputs for the neural network architecture with a CNN layer that predicts signals. In their study, they examined and tested the structure of the English alphabet from A to Z and claimed to achieve 100% recognition rate and 48% sound rejection of static signs/letters. For gesture recognition, it comes first by converting the RGB image to YUQ or YIQ color space for edge segmentation and skin segmentation to determine the edge of the hand [7]. Then use the convex body method to manually check for fingerprints. Finally, neural networks are used as a classification method. The model eventually achieved an accuracy of 98.2%. Sharma et al. [8] developed a system based on Indian Sign Language for communication with the speech or hearing impaired. First, after the image is captured, the data is preprocessed by converting it from RGB to grayscale using the Matlab environment [8]. Then use the Sobel test to identify the edges of the image using a 3×3 filter. Finally, the hierarchical centering method was applied to the 600-point subimage; With this method, 124 features were obtained. The classification techniques used are KNN and Neural Networks. The accuracy obtained with this method is 97.10%. Agarwal et al. [9] By using a glove sensor to sign, make symbols, and present content in parentheses, it aims to bridge the gap between people with speech impairments and those with normal speech skills. Subjects gestured using the glove sensor [9].

III. METHODOLOGY

Building SignAssist using deep learning involves collecting a diverse dataset, preprocessing the data, feature extraction, model training, and model evaluation. OpenCV, TensorFlow, and Keras are powerful tools that simplify the implementation of these steps. The resulting model can be deployed as a standalone application or integrated into existing applications to provide accessibility to sign language users.

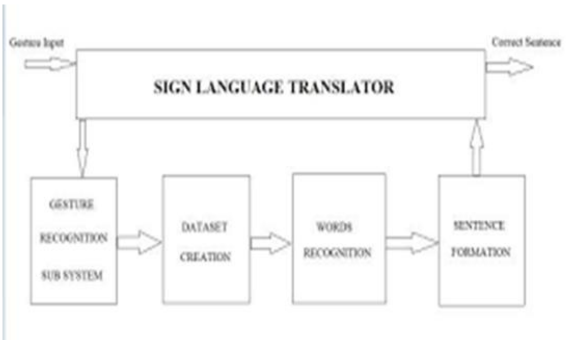


Fig 1: Block Diagram of System

Fig 2: Sign Language Interpreter Flowchart

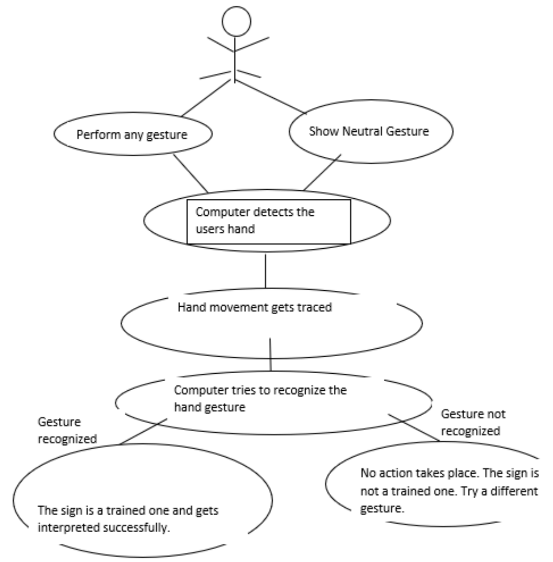


Figure 3 Use Case Diagram

A. Dataset

Kaggle's American Sign Language dataset has been used to train the model.

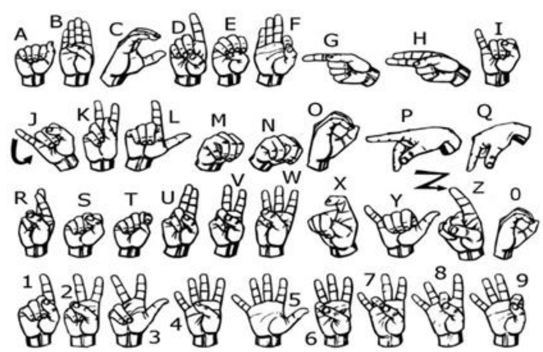


Fig 4: American Sign Language Alphabets and Digits

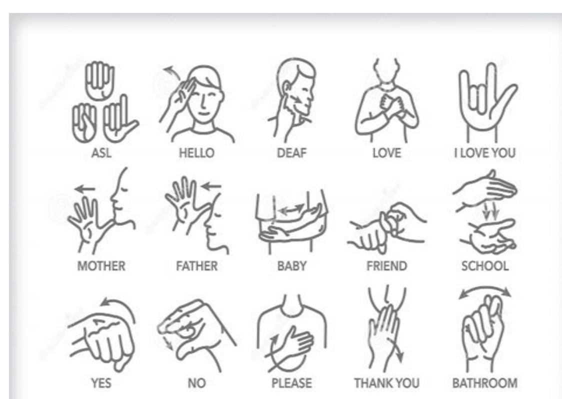


Fig 5: Common Words in American Sign Language

IV. IMPLEMENTATION

In the proposed system, TensorFlow, Keras, and OpenCV have been used to further the advancement in interactions between humans and machines. The proposed system translates American Sign Language for sign language users by using a mechanism that utilizes the webcam and live video stream of the computer.

To implement an accurate and efficient American Sign Language interpreter using deep learning, the system focuses on a comprehensive methodology. It begins by carefully selecting a suitable dataset for training and evaluation. In this study, Kaggle's dataset of sign language videos has been utilized, which provides a diverse range of gestures and corresponding sign language glosses.

In order to prepare the data for deep learning, we employed OpenCV, a powerful computer vision library. Through OpenCV, we performed essential preprocessing steps to enhance the quality and suitability of the dataset. Specifically, we segmented the sign language videos into individual frames, resized them to a consistent size, and converted them to grayscale. By converting the frames to grayscale, we reduced the dimensionality of the data while retaining crucial information regarding handshapes and movements.

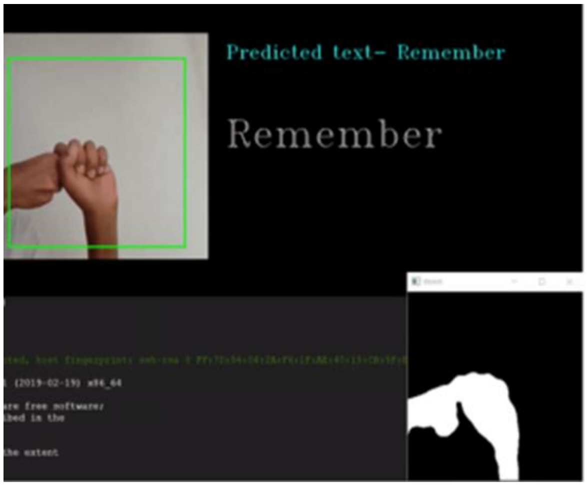
To capture the intricate details of sign language, we utilized a deep learning model architecture consisting of convolutional neural network (CNN) components. TensorFlow and Keras, well-known frameworks for deep learning, enabled us to build and train this sophisticated model. The CNN component facilitated the extraction of spatial features from the preprocessed video frames, while the subsequent recurrent layers captured the temporal dependencies inherent in sign language gestures.

To evaluate the performance of our model, we meticulously split the dataset into separate training and validation sets. This allowed us to train the model on a substantial portion of the data while reserving a subset for evaluating its accuracy and generalization capabilities. During training, we employed appropriate optimization techniques, such as stochastic gradient descent, to iteratively update the model's weights and biases, optimizing its ability to recognize sign language gestures accurately.

Upon achieving a satisfactory level of performance, we deployed the trained model as a standalone application using the Python programming language. This application harnessed users' webcams, enabling them to capture real-time sign language gestures. The model then classified these gestures and provided corresponding sign language interpretations, offering a practical and accessible means of communication for individuals interacting with the application.

Throughout the implementation process, we remained attentive to user feedback and continuously sought to enhance the interpreter's accuracy and usability. By incorporating user feedback, we addressed potential limitations, refined the model's predictions, and further optimized the application's performance.

In conclusion, our research paper presents a robust implementation of an American Sign Language interpreter using deep learning techniques. By collecting and annotating a diverse dataset, leveraging OpenCV for preprocessing, utilizing TensorFlow and Keras for model building and training, and deploying the model as a real-time application, we achieved a practical and efficient solution for facilitating communication between sign language users and non-signing individuals. User feedback was gathered to improve the accuracy and usability of the application. Based on the feedback, we will be looking to make several improvements, including adding support for more sign language gestures. Some sign interpretations of the proposed model are:



V. APPLICATIONS

SignAssist is helpful in many ways: moving the cursor throughout the screen takes physical effort and elderly people, children, and people with disabilities often find it difficult to control the computer by themselves and are dependent on others to do the job for them. Therefore, the proposed system has got them covered as it greatly improves the ease of usability of computers.

The major applications of the proposed system are: (i)It can be used to interact with machines by people who use sign language(ii)It can be used to communicate with the computer virtually which reduces physical and cognitive effort(iii)With the help of this, sign language users do not have to be dependent on interpreters(iv)Users can efficiently convey their symptoms to a doctor(v)It enables them to ask questions during educational sessions and communicate effectively in various settings.

CONCLUSION

The objective of our proposed system is to enable efficient communication between sign language users and machines This model can be achieved with the help of an external or inbuilt webcam that processes a live video feed and recognizes hand signs to identify and interpret specific commands/dialogues.

Our proposed system has an accuracy of >95%, which is far greater than that of pre-existing systems, and it has many applications. Based on experimental results, we can conclude that the proposed Deep Learning based Sign Language Interpreter system has very good performance and high accuracy. Since the proposed model is more accurate, it is implementable in real life.

This model has some limitations, such as a slightly limited vocabulary pool and a lack of feedback mechanism. Also, it only interprets American Sign Language. We aim to address the limitations of our current approach by developing the system further to improve its functioning and execution. Further, we will try to add more features to this application for better functionalities which will scale it further and help reach a better audience.

Thus, SignAssist is a wonderful mechanism for sign language users. It makes interactions easier as it makes sign language users independent of interpreters, giving the users a delightful experience overall.

VI. FUTURE SCOPE

Limitations of the proposed system are that the vocabulary pool is limited and there is no feedback mechanism. Also, the system can only interpret American Sign Language.

To further the scope of this system, we can increase the vocabulary pool of our model. Further, we can incorporate a feedback mechanism to make the system more robust. We can also train it to interpret more sign languages.

REFERENCES

[1] Chen, J. CS231A Course Project Final Report Sign Language Recognition with Unsupervised Feature Learning 2012. Available online: http://vision.stanford.edu/teaching/cs231a_autumn1213_internal/project/final/writeup/distributable/Chen_Paper.pdf (accessed on 15 March 2022).

[2] Deepsign: Sign Language Detection and Recognition Using Deep Learning Deep Kothadiya, Chintan Bhatt, Krenil Sapariya, Kevin Patel, Ana-Belén Gil-González, and Juan M. Corchado

[3] Sign Language Recognition using Deep Learning, Authors: Dhruv Sood, https://www.ijraset.com/research-paper/sign-language-recognition-using-deep-learning

[4] Adaloglou, N.; Chatzis, T. A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition. IEEE Trans.Multimed. 2022,24, 1750–1762.

[5] Jiang, S.; Sun, B.; Wang, L.; Bai, Y.; Li, K.; Fu, Y. Skeleton aware multi-modal sign language recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 21–24 June 2021; pp. 3413–3423.

[6] Wadhawan, A.; Kumar, P. Sign language recognition systems: A decade systematic literature review. Arch. Comput. Methods Eng.2021,28, 785–813.

[7] Yang, D.; Martinez, C.; Visuña, L.; Khandhar, H.; Bhatt, C.; Carretero, J. Detection and Analysis of COVID-19 in medical imagesusing deep learning techniques. Sci. Rep. 2021,11, 19638.

[8] Likhar, P.; Bhagat, N.K.; Rathna, G.N. Deep Learning Methods for Indian Sign Language Recognition. In Proceedings of the 2020IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), Berlin, Germany, 9–11 November 2020; pp. 1–6.

[9] De Coster, M.; Herreweghe, M.V.; Dambre, J. Sign Language Recognition with Transformer Networks. In Proceedings of the Conference on Language Resources and Evaluation (LREC 2020), Marseille, France, 13–15 May 2020; pp. 6018–6024.

[10] Aparna, C.; Geetha, M. CNN and Stacked LSTM Model for Indian Sign Language Recognition. Commun. Comput. Inf. Sci.2020,1203, 126–134.

[11] Mittal, A.; Kumar, P.; Roy, P.P.; Balasubramanian, R.; Chaudhuri, B.B. A Modified LSTM Model for Continuous Sign LanguageRecognition Using Leap Motion. IEEE Sens. J. 2019,19, 7056–7063.

[12] Yan, S. Understanding LSTM and Its Diagrams. Available online: https://medium.com/mlreview/understanding-lstm- and-its-diagrams-37e2f46f1714

[13] Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections onLearning. arXiv 2016, arXiv:1602.07261.

[14] Muthu Mariappan, H.; Gomathi, V. Real-Time Recognition of Indian Sign Language. In Proceedings of the International Conference on Computational Intelligence in Data Science, Haryana, India, 6–7 September 2019.14.Mittal, A.; Kumar, P.; Roy, P.P.; Balasubramanian, R.; Chaudhuri, B.B. A Modified LSTM Model for Continuous Sign LanguageRecognition Using Leap Motion. IEEE Sens. J. 2019,19, 7056–7063.

[15] Camgoz, N.C.; Hadfield, S.; Koller, O.; Ney, H.; Bowden, R. Neural Sign Language Translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018, Salt Lake City, UT, USA, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018.

[16] Shivashankara, S.; Srinath, S. American Sign Language Recognition System: An Optimal Approach. Int. J. Image Graph. SignalProcess. 2018,10, 18–30

[17] Wazalwar, S.S.; Shrawankar, U. Interpretation of sign language into English using NLP techniques. J. Inf. Optim. Sci.2017,38, 895–910.

[18] Manware, A.; Raj, R.; Kumar, A.; Pawar, T. Smart Gloves as a Communication Tool for the Speech Impaired and Hearing Impaired.Int. J. Emerg. Technol. Innov. Res. 2017,4, 78–82.

[19] V. Bheda and D. Radpour, "Using deep convolutional networks for gesture recognition in American sign language," arXiv:1710.06836, 2017.

[20] Sharma, M.; Pal, R.; Sahoo, A. Indian sign language recognition using neural networks and KNN classifiers. J. Eng. Appl. Sci.2014,9, 1255–1259.

[21] Thomas, J.; McDonagh, D. Shared language: Towards more effective communication. Australas. Med. J. 2013, 6, 46–54

[22] Chen, J.K. Sign Language Recognition with Unsupervised Feature Learning; CS229 Project Final Report; Stanford University: Stanford, CA, USA, 2011.

[23] Mekala, P.; Gao, Y.; Fan, J.; Davari, A. Real-time sign language recognition based on neural network architecture. In Proceedings of the IEEE 43rd Southeastern Symposium on System Theory, Auburn, AL, USA, 14–16 March 2011.

[24] Nandy, A.; Prasad, J.; Mondal, S.; Chakraborty, P.; Nandi, G. Recognition of Isolated Indian Sign Language Gesture in Real Time.Commun. Comput. Inf. Sci. 2010,70, 102–107.

[25] Efthimiou, E.; Fotinea, S.E.; Vogler, C.; Hanke, T.; Glauert, J.; Bowden, R.; Braffort, A.; Collet, C.; Maragos, P.; Segouat, J. Sign

[26] Language Recognition, Generation, and Modelling: A Research Effort with Applications in Deaf Communication. In UAHCI

[27] 2009: Universal Access in Human-Computer Interaction. Addressing Diversity; Lecture Notes in Computer Science; Stephanidis, C., Ed.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5614._3.

[28] Gao, W.; Fang, G.; Zhao, D.; Chen, Y. A Chinese sign language recognition system based on SOFM/SRN/HMM. Pattern Recognit. 2004, 37, 2389–2402.

[29] Liang, R.-H.; Ouhyoung, M.I. A Sign Language Recognition System Using Hidden Markov Model and Context Sensitive Search. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology, Hong Kong, China, 1–4 July 1996.

[30] Chiarelli, B. The Origin of Human Language, Studies in Language Origins; Publisher John Benjamins: Amsterdam, The Netherlands,1991; 35p.