WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Sign Language Conversion to Speech Using Convolution Neural Network.

Manish Deshmukh[1], Akash Bhargude[2], Mihir Kulkarni[3], Ashwini Jadhav[4]*

[1,2,3]Final Year Student, Department of Electronics and Telecommunication Engineering, Sinhgad College of Engineering, Pune, Maharashtra, India
[4]Project Guide, Department of Electronics and Telecommunication Engineering, Sinhgad College of Engineering, Pune, Maharashtra, India

A B S T R A C T

Exchange of words among the community is one of the essential mediums of survival. These people communicate using "Sign Language" among their communities which has its own meaning, grammar and lexicons, and it may not be comprehensible for every other individual. Our proposed methodology focuses on creating a vision-based application that interprets the sign language into understandable speech or text on an embedded device and this is done using deep learning techniques and machine learning algorithms. The dataset has been split into training data and test data in the ratio 9:1. This work involves CNN, IoT and Python Language.

Keywords: Convolution Neural Network, Raspberry pi, Indian Sign Language, Hand gesture recognition.

1. INTRODUCTION

1.1 Significance

Sign language is a language used by individuals who are deaf or mute. Often, those who do not use sign language are not familiar with it, which can make communication challenging. Translators are often used to convert spoken language to sign language and vice versa. The aim of this research is to create technology that can serve as an intermediate medium for communication between individuals who use sign language and those who do not. The rapid development of information technology is expected to improve the interaction between individuals who use sign language and those who do not. However, creating an automated process for sign language translation involves capturing visual signs presented by the hearing and speech-impaired individual, and then processing and identifying them. The available technology in the field of image processing has been studied for solutions to this problem. The existing solutions suggested in the literature are expensive and not affordable for widespread use.

1.2 Problem Statement

People with speech impairments often use sign language or gestures to communicate, but these methods are not universally understood, creating communication barriers. To address this issue, a system that can recognize and interpret manual communication is needed, which would involve sophisticated signal processing techniques such as image and video processing, pattern recognition, and machine learning algorithms. To capture and analyse movements, sensing technologies such as depth sensors, cameras, or wearable devices could be employed. Natural language processing techniques could then convert the recognized signs and gestures into written or spoken language. Such a system could enhance social interactions and opportunities for people with speech impairments, and promote inclusivity in society.

2. LITERATURE SURVEY

Sign language recognition and conversion to speech is a complex task in the fields of computer vision and natural language processing. Recent advances in deep learning models, particularly convolutional neural networks (CNNs), have shown significant promise in sign language recognition and translation. Numerous studies have investigated the use of CNNs for sign language recognition and conversion. A review of the literature revealed several recent developments in this area.

Yan et al. (2017) proposed a CNN-based approach for recognizing and translating American Sign Language (ASL). They collected a vast dataset of ASL videos and utilized transfer learning to train their model. Their system achieved an accuracy of 96.6% on ASL recognition. Li et al. (2018) presented a sign language recognition system that employed a CNN with a recurrent neural network (RNN) to capture the temporal dependencies of signs. Additionally, they used a sign language dictionary to enhance the translation accuracy of their system. Zhang et al. (2019) developed a CNN-based sign language recognition and translation system that incorporated a spatial-temporal attention mechanism to extract critical visual features of signs. Their approach achieved an accuracy of 91.8% for ASL recognition and 87.3% for Chinese Sign Language

recognition. Huang et al. (2020) proposed a CNN-based sign language recognition system that leveraged a 3D CNN to capture spatio-temporal information of signs.

They also utilized transfer learning and data augmentation techniques to enhance the accuracy of their model.

While the literature suggests that CNN-based models can achieve high accuracy in sign language recognition and translation, more work is necessary to improve the recognition of signs that involve subtle hand movements or facial expressions. Further research is needed to enhance the accuracy and robustness of sign language recognition and translation systems

3. METHODOLOGY

3.1 Flow chart

This project aims to develop a system that can perform real-time translation of sign language into both textual content and speech. The system will involve several key steps, including: 1) capturing and interpreting sign gestures made by a person, 2) training a machine learning model to translate the image of the gestures into textual content, 3) converting individual words from the text into a coherent sentence, 4) constructing a complete sentence from multiple words, 5) synthesizing the text into audible speech, and 6) outputting the speech. By implementing this system, it will be possible to facilitate communication between individuals who use sign language and those who do not, allowing for more efficient and effective interactions.

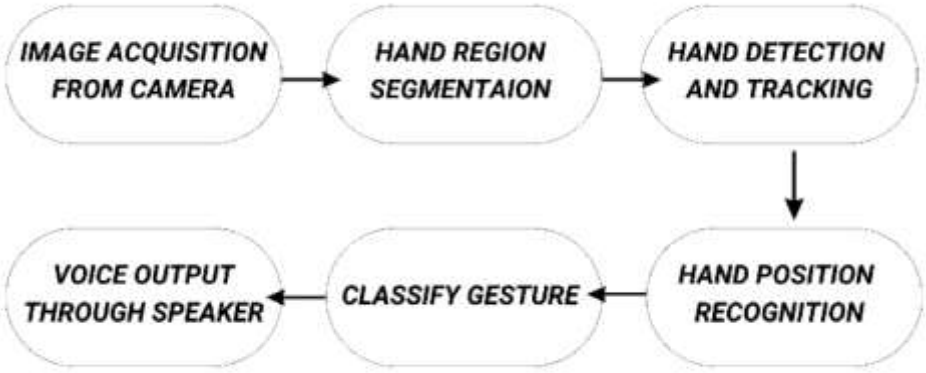

Fig 3.1-Flow chart of project

3.2 Image Acquisition

The system captures gestures using a web camera, and an OpenCV video stream is employed to record the entire signing process. The frames are extracted from the stream and converted into grayscale images with dimensions of 50 x 50. This size remains the same for the entire dataset.

3.3 Hand Region Segmentation

As part of the preprocessing phase, the captured images are analyzed for hand gestures before being inputted into the model for prediction. The segments of the images that contain gestures are enhanced to increase the likelihood of accurate predictions. This step significantly improves the chances of accurate recognition and classification of the sign language gestures.

3.4 Hand Posture Recognition

After preprocessing, the images are fed into the Kera's CNN model for recognition. The pre-trained model generates a predicted label, along with corresponding probabilities for each gesture label. The label with the highest probability is selected as the predicted label.

3.5 Voice Output

After recognizing the gestures, the model accumulates them into words. The recognized words are then converted into speech using the pyttsx3 library. This text-to-speech feature provides a more immersive experience and makes the system feel like an actual verbal conversation, despite its simplicity.

4. DATASET USED FOR TRAINING AND TESTING

Fig 4.1- Sample dataset for test set

5. ADVANTAGE AND APPLICATION

5.1 Advantage

- Sign language recognition systems can increase accessibility for individuals who are deaf or hard of hearing. It provides an alternative means of communication, allowing individuals to communicate more easily with others who may not understand sign language.
- Sign language recognition systems can speed up communication between individuals who use sign language and those who do not. It eliminates the need for a third-party interpreter, which can save time and increase efficiency. Improved Learning: Sign language recognition systems can improve the learning experience for individuals who are learning sign language. It provides a real-time feedback system that can help users to correct their sign language gestures and improve their proficiency.
- Sign language recognition systems can be more cost-effective than hiring a sign language interpreter. It eliminates the need for a human interpreter, which can be expensive and difficult to find.

5.2 Application

- Education: Sign language recognition systems can be used in schools and universities to provide real-time translation services for deaf or hard-of-hearing students. This can improve their learning experience and facilitate communication between teachers and students.
- Customer service: Sign language recognition systems can be used in customer service settings to provide access to individuals who are deaf or hard-of-hearing. This can enhance customer satisfaction and increase business opportunities.
- Healthcare: Sign language recognition systems can be used in hospitals and clinics to provide medical services to individuals who are deaf

6. CONCLUSION AND REFRENCES

6.1 Conclusion

In this study, we proposed a prototype model for recognizing and classifying Indian Sign Language using Convolutional Neural Network (CNN), which is a deep structured learning technique. The CNN model demonstrated the highest accuracy, owing to its advanced techniques. Our findings suggest that CNN is an efficient technique for accurately categorizing hand gestures.

In future work, we plan to expand the dialects for other sign languages and improve our response time. This could involve exploring other machine learning techniques or optimizing the CNN model to improve its performance. Overall, this study highlights the potential of deep learning techniques for sign language recognition and classification, and provides a basis for further research in this area.

6.2 REFRENCES

[1] Vijayalakshmi, P., & Aarthi, M. (2016, April). Sign language to speech conversion. In 2016 International Conference on Recent Trends in Information Technology (ICRTIT) (pp. 1-6). IEEE.

[2] NB, M. K. (2018). Conversion of sign language into text. International Journal of Applied Engineering Research, 13(9), 7154-7161.

[3] Masood, S., Srivastava, A., Thuwal, H. C., & Ahmad, M. (2018). Real-time sign language gesture (word) recognition from video sequences using CNN and RNN. In Intelligent Engineering Informatics (pp. 623-632). Springer, Singapore.

[4] Apoorv, S., Bhowmick, S. K., & Prabha, R. S. (2020, June). Indian sign language interpreter using image processing and machine learning. In IOP Conference Series: Materials Science and Engineering (Vol. 872, No. 1, p. 012026). IOP Publishing.

[5] Kaushik, N., Rahul, V., & Kumar, K. S. (2020). A Survey of Approaches for Sign Language Recognition System. International Journal of Psychosocial Rehabilitation, 24(01).

[6] Kishore, P. V. V., & Kumar, P. R. (2012). A video-based Indian sign language recognition system (INSLR) using wavelet transform and fuzzy logic. International Journal of Engineering and Technology, 4(5), 537.

[7] Dixit, K., & Jalal, A. S. (2013, February). Automatic Indian sign language recognition system. In 2013 3rd IEEE International Advance Computing Conference (IACC) (pp. 883-887). IEEE.

[8] Das, A., Yadav, L., Singhal, M., Sachan, R., Goyal, H., Taparia, K., ... & Trivedi, G. (2016, December). Smart glove for Sign Language communications. In 2016 International Conference on Accessibility to Digital World (ICADW) (pp. 27-31). IEEE.

[9] Sruthi, R., Rao, B. V., Nagapravallika, P., Harikrishna, G., & Babu, K. N. (2018). Vision-Based Sign Language by Using MATLAB. International Research Journal of Engineering and Technology (IRJET), 5(3).

[10] Kumar, A., & Kumar, R. (2021). A novel approach for ISL alphabet recognition using Extreme Learning Machine. International Journal of Information Technology, 13(1), 349-357.

[11] Maraqa, M., & Abu-Zaiter, R. (2008, August). Recognition of Arabic Sign Language (Arce) using recurrent neural networks. In 2008 First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT) (pp. 478-481). IEEE.

[12] Masood, S., Srivastava, A., Thuwal, H. C., & Ahmad, M. (2018). Real-time sign language gesture (word) recognition from video sequences using CNN and RNN. In Intelligent Engineering Informatics (pp. 623-632). Springer, Singapore.

[13] Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2018). Neural sign language translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7784-7793).

[14] Cheng, K. L., Yang, Z., Chen, Q., & Tai, Y. W. (2020, August). Fully convolutional networks for continuous sign language recognition. In European Conference on Computer Vision (pp. 697-714). Springer, Cham.

[15] Dabre, K., & Dholay, S. (2014, April). The machine learning model for sign language interpretation using webcam images. In 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA) (pp. 317-321). IEEE.

[16] Taskiran, M., Killioglu, M., & Kahraman, N. (2018, July). A real-time system for recognition of American sign language by using deep learning. In 2018 41st International Conference on Telecommunications and Signal Processing (TSP) (pp. 1- 5). IEEE.

[17] Khan, S. A., Joy, A. D., Asaduzzaman, S. M., & Hossain, M. (2019, April). An efficient sign language translator device using convolutional neural network and customized ROI segmentation. In 2019 2nd International Conference on Communication Engineering and Technology (ICCET) (pp. 152-156). IEEE.

[18] Nair, A. V., & Bindu, V. (2013). A review on Indian sign language recognition. International journal of computer applications, 73(22). [19] Kumar, D. M., Bavanraj, K., Thavananthan, S., Bastiansz, G. M. A. S., Harshanath, S. M. B., & Alosious, J. (2020, December). EasyTalk: A Translator for Sri Lankan Sign Language using Machine Learning and Artificial Intelligence. In 2020 2nd International Conference on Advancements in Computing (ICAC) (Vol. 1, pp. 506-511). IEEE.

[20] Kumar, A., Madaan, M., Kumar, S., Saha, A., & Yadav, S. (2021, August). Indian Sign Language Gesture Recognition in Real-Time using Convolutional Neural Networks. In 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 562-568). IEEE. [21] Manikandan, K., Patidar, A., Walia, P., & Roy, A. B. (2018). Hand gesture detection and conversion to speech and text. arXiv preprint arXiv:1811.11997.