Sign Language Translation to Text Using Deep Learning

Dr. G. Niranjana
SRM UNIVERSITY OF SCIENCE AND TECHNOLOGY
Kattankulathur, Chennai , India

SHIVEN PAUDYAL
B.TECH(Computer science and engineering)
SRM UNIVERSITY OF SCIENCE AND TECHNOLOGY
Kattankulathur, Chennai , India
sp6691@srmist.edu.in

SAHIL ALI
B.TECH(Computer science and engineering)
SRM UNIVERSITY OF SCIENCE AND TECHNOLOGY
Kattankulathur, Chennai , India
sa8031@srmist.edu.in

Abstract— Sign language is a way of communication using hand gestures instead of words that is mainly used by people who are deaf or have hearing impairments. As most of the people are unaware of the sign language gestures, we intend to build a real time sign language to text converter using CNN. We will make use of fingerspelling which is the representation of the letters of a writing system that only makes use of hand gestures. The captured image of the hand will be initially passed through a filter followed by the application of a classifier that will predict the class of the hand gesture thereby predicting the letter and word. By building this model we intend to bridge the gap between those who can and cannot interpret sign language, ergo, promoting unity among all. In order to overcome many obstacles and improve people's daily lives, understanding hand signals is crucial. There are many potential uses for machines that can decipher human activity and its meaning. The identification of sign language is one area garnering particular attention. In-depth analysis of current advances in sign language and hand gesture detection technologies is provided in this paper. The five procedures of data gathering, pre-processing, segmentation, feature extraction, and classification are properly employed to divide the approaches under consideration. Alternative algorithms are described at each level, and their benefits are compared. We also discuss the challenges and limitations that research on general gesture recognition and specifically sign language understanding encounter.

I. INTRODUCTION

Sign language is essential for deaf people to communicate with others and themselves, but ordinary people pay little attention to sign language. When we don't have deaf relatives, we as ordinary people tend to overlook the value of sign language.
Using the services of a sign language interpreter is a great way to communicate with people who are deaf or mute. However, hiring of such kind of personnel is very expensive. Acost-effective solution is needed to enable deaf and dumb

people to have normal conversations.
Scientists are therefore working to develop means of communication for deaf people to interact with deaf people. The sign language recognition system is an innovation. This technology attempts to understand sign language and translate it into your native language, either orally or in text form. However, this technology was expensive to develop and difficultto use daily. Early research shows that data gloves can be used to recognize sign language. However, the high cost of gloves and the nature of wearing them make them difficult to commercialize. Researchers then attempt to take this knowledge into account to create a purely visual sign language recognition system. However, it also has its drawbacks, especially when trying to track hand movements properly.
From taking pictures to classifying them, problems can arise in the development of sign language recognition. The ideal technique for capturing images is still being researched by scientists. The challenge of image preprocessing is illustrated by image acquisition using a camera. However, using active sensor devices can be costly. The use of classification systems by researchers also has certain drawbacks. With so many possibilities, researchers are unable to determine the best detection method. Focusing on one method tends to avoid testing other methods that are more suitable for sign language recognition. Researchers rarely attempt other approaches to fully formulate strategies.
The purpose of this essay is to discuss sign language recognition systems used by scientists. This white paper describes an application view of sign language recognition. This essay discusses the devices used to collect the data, the data collection, including data from early studies or self-generated data, detection methods recently used by researchers, and the results of previous studies.

II. METHOD

The next few or the following are based on the internet data/databases we used/utilized in our research: Wiley online library, JIIP, MDPI, IEEE, IRJET etc.

Procedure which was followed while seeking publication in the online databases:

1. First, we must type a term in the online databases.
2. Now filter it to desire timeline which should 2017-2021.
3. Sort it to studies into the discovered studies document.
4. Now sort the previous studies into the filtered studies document which is based on title, abstract, and result.
5. Choose only the relevant content to put in your study.

Table 1. Sign Language Translation

S.NO	Paper Name:	Author Name:	Journal :	Publication Year:	Summary
1.	"Indian Sign Language Recognition Using Machine" [1]	"Ashok Kumar Sahoo"	Wiley Online Library	17 June 2021	A sign database with ISL digits has been constructed that has 5000 photos, 500 images for each numerical sign, in order to recognize ISL sign images in a real-time context.
2.	"Real Time Sign Language Recognition and Speech" [2]	"Amrita Thakur", "PuianBudhathoki", "Sarmila Upreti", "Shirish Shrestha", "Subama Shakya"	Journal of Innovative Image Processing(JIIP)	Feb 2020	The paper focuses on model training over ASL dataset. Model used is VGG-16.
3.	"Sign Language Detection and Recognition Using Deep Learning" [3]	"Deep Kothadiva", "Chintan Bhatt", "KrenilSapativa", "Kevin Patel", "Ana-Belén Gil-Gonzalez" and "Juan M. Corchado"	MDPI	3 June 2022	In this paper, the authors divided the video files containing the ISL into sub section videos, so that each file contains one word only.

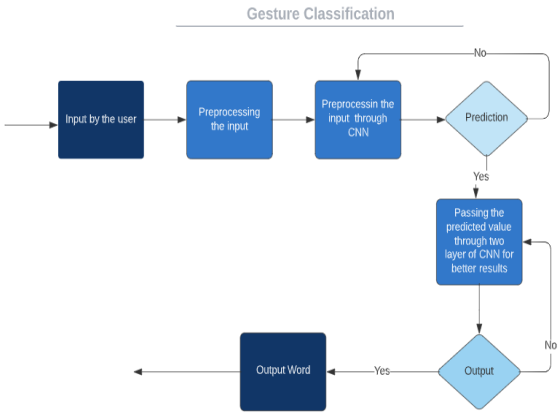
4.	"Sign Language Recognition UsingDeep Learning" [4]	"Aditya Das Shantanu gawde", "Khyati. Suratwala", "Dhananjay Kalbande"	IEEE	2018	Uses custom Processed Static Gesture Images
5.	"Real Time sign language using deep learning" [5]	"Sanket Bankar.Tushar Kadam" "Vedant Korhale", "Mrs. A.A. Kulkarni"	IRJET	APRIL.2022	This paper successfully worked on the limitations of CNN during livestream, which is CNN [21] is slow andthe sign needs to be in a constrained area to get detected.
6.	"Sign Language Recognition System using TensorFlow" [6]	Sharvani Srivastava, Amisha Gangwar Richa Mishra, SudhakarSingh	University ofAllahabad, Prayagraj	NA -2021	"The model used is SSD MobileN etv2 320x320 from tensortiow, modelzoo which is present in TensorFlow, object detection API".
7.	"Sign language detection using deeplearning" [7]	"Ekta Sood", "Simon Tannert", "Diego Frassinelli", "Andreas Bulling", "Ngoc Thang Vu"	URASETI	NA - 2021	"A set of fully connected layers is created and added to them after the inception network to tailorthe neural network".

8.	"Sign Language Recognition Systemusing Convolutional Neural Network and Computer Vision"[8]	Mehreen Hurroo	IJERT	11-12-2020	"The photos go through anumber of processes of processing, including several Computer vision techniques including grayscale conversion, dilatation, and mask operation. Additionally,the area of attention, in this example the hand gesture, is split".
9.	"Sign talk: Sign Language to Text andSpeech Conversion C.Uma Bharathi" [9]	"G. Ragavi;K.Karthika"	IEEE	2021	"The proposed work makesit easier for people who have trouble hearing, speaking, or even seeing to interact with others".
10.	"A Report on Translating Sign Language toEnglish Language "[10]	"Mahender Reddy Chilukala", "Vishwa Vadalia"	IEEE	2022	"To identify signs from diverse languages, such as American Sign Language, Indian Sign Language, and Indonesian Sign Language, analysis of various approaches had tobe done".
11.	"Sign Language toTextfor Deafand Dumb" [11]	"Vibhu Gupta", "Mansi Jain", "GarimaAggarwal"	IEEE	2022	"The experimental findings demonstrate that this work outperformed most recent research publications on American gesture-based communication, with an accuracy score of 99.8%on testing data".

12.	"Sign Language toText Conversion using Deep Learning" [12]	"Aruna Bhat", "Vinay Yadav", "Vishesh Dargan", "Yash"	INCET, IEEE	2022	"As part of this project, a variety of images with American Sign Language signs have been processedand used to implement a CNN [21] model that will classify the images as their alphabetical Equivalents".
13.	"An Efficient Sign Language Translator Device Using Convolutional NeuralNetwork and Customized ROI Segmentation" [13]	"Saleh Ahmad Khan;Amit Debnath Joy; S. M. Asaduzzaman;Morsali nHossain"	ICCET	2019	"This technique offers an effective model that, in the end, makes it simple to add more indicators to the final Raspberry Pi prototype".
14.	"Real-time American SignLanguage Recognition with Convolutional Neural Networks" [14]	Brandon Garcia	Stanford University	2015	"The development of a real-time sign language interpreter is a significantstep in improving communication between the deaf population and the general public. The creation and application of an ASL fingerspelling translator based on a CNN [21]are presented here".
15.	"Dynamic sign language translating system usingdeep learningand natural language processing" [15]	"Aishwarya Kulkarni", "Pranav Halgekarb", "Girish.R.Deshpande", "Anagha Raod", "Aishwarya Dinn"	Turkish Journal of Computer and Mathematics Education	28 April 2021	"Natural language processing [23], text-to- speech algorithms, Convolutional neural networks, and language translation are all used bythe system. It is able to accurately forecast the letters needed to make the intended sentence from hand gestures and

					recognise them dynamicall y".
16.	"Indian Sign Language Recognition using ConvolutionalNe ural Network" [16]	"Rachana Patil"	ICACC	2021	"Our model is trained using a convolutional neural network (CNN) [21], which is then utilisedto recognise the images. Our model has a 95% accuracy rate".
17.	"Conversion ofSign Language Video to Text and Speech" [17]	"Panuganti Usha sri"	IJRASET	2022-04-30	"When a video is providedas input, our method of sign language translation uses CNN [21] and RNN [26] to identify the signs and translate them intotext and speech".
18.	"Speech To Sign Language Conversionusing Convolutional Neural Networks" [18]	"Sreeraksha M R", "Vani H Y", "PhaniBhushan", "D K Shivkumar"	IJERT	20-07-2021	"Using MFCC [27] as a feature extraction methodand CNN [21] as a classifier, the suggested method has a 93%accuracy rate".
19.	"Sign Language to Text Conversion –A Survey" [19]	"Adithi Krishna", "Ruthvik B R", 'Spoorthy M", "Rhea Muthanna","Shashank"	InternationalJournal of Scientific & Engineering Research	2019	"Our major goal is to create a smartphone application that translatessign language into text while also making some effort to somewhat address the aforementioned issues".
20.	"Arabic Sign Language Recognition and Generating Arabic SpeechUsing Convolutional NeuralNetwork" [20]	"Yin Zhang"	M. M. Kamruzzaman	23 May 2020	"Arabic hand sign-based letters are recognised, andthe results are sent into the text-to-speech engine,which outputs audio of the Arabic language".

III. PROPOSED METHODOLGY



The above figure shows us the best way to convert a sign language totext or speech is by using CTC to separate words at best to build asentence or sentence level structure of translation. But before that we should use CNN model for classification which helps us to recognize the symbol of any type of sign language be it ASL, ISL any. Convolutional layers, which subject the images to a predetermined number of convolution filters. The layer makes several calculations on each subregion to create a single value for the output feature map. In order to inject nonlinearities into the model, convolutional layers usually utilise a ReLU activation function to the output. dense layers that classify the features that have been retrieved and down-sampled by the pooling and convolutional layers. The nodes in a dense layer are all linked to those in the layer underneath it.

IV. IMPLEMENTATION

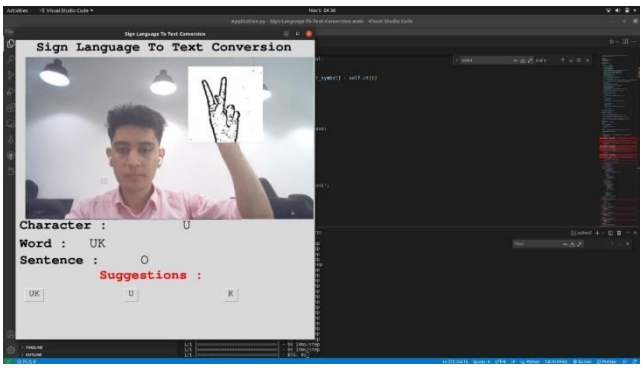
In this we will get dataset by using OpenCV where we will get live images. The dataset comprises of images of different hand signs based on the American sign language (ASL). Then we will create a Region of interest (ROI) to create a part (specifically box) in the screen where our hand signs will be captured. Now we will capture the image and load it into dataset for the model's Training. For better performance of the model, we will take the picture of the same gesture in many ways so that it will be easy for the model to predict even if the hand sign is not that clear. After capturing images, we will train our dataset using the data we saved and find its accuracy. We compute the background's cumulative weighted average to differentiate it from the foreground by subtracting it from the frames that have a distinguishable foreground item in front of the background. After the training we will optimize our model's accuracy using keras and modifying the hyperparameter (if error occurs). Keras is used for the implementation of neural networks and it makes it easy. For the model's Accuracy we will also try different algorithms and find out which algorithm gives accuracy and validation the most. Convolutional neural networks (CNN) were utilized in this instance to generate the model. A convolutional neural network in contrast to traditional neural networks, CNN layers arrange the neurons in three dimensions: width, height, and depth. The neurons in a layer will only be connected to a tiny piece of the layer (window size) preceding it in order to compress the entire image into a single vector of class scores at the conclusion of the CNN architecture. In addition to dimensions (number of classes), the final output layer will also have neurons that are not fully linked.
The architecture of a CNN consists of multiple convolution layers, each of which includes a pooling layer, activation function, and optional batch normalization, as well as fully connected layers. As the image passes through the CNN layers, its size is reduced due to max pooling. The last layer provides the predicted class probabilities.

V. CONCLUSION

After completing and making the changes in various sectors of our code we are finally able to predict the hand sign gesture that has been shown in the region of Interest and given the constraints it has performed exceptionally well. The images that were loaded in the dataset predict the letters under the right condition, that is proper light, solid background and no hindrance while predicting the gesture. As we know nothing is perfect, everything needs improvement and so is the case with our project, there are few hardware problems that can hinder the results of our prediction.
There is always room for improvement, So in our project there can be few enhancements that can be done with proper facilities, that is better Random Access Memory (RAM), More Computational Power, Higher Processor and to create a dataset that can be done by only by providing multiple data from various people.

VI. RESULTS

For the model's Accuracy we will also change some algorithms and find out which algorithm gives accuracy and validation the most. We have used Convolutional Neural Network and added extra layers so that the prediction becomes accurate and the chances of predicting the right gesture increase. The additional layer apart from the first two serves the purpose of increasing the accuracy of the model. By doing so, the time for prediction is also increased a little but time taken to predict gestures by a system depends on the system's capacity and hardware component. The final output is shown by the following Image where a pop up window appears, comprising of a white and black rectangular area meant for removing background noises and getting proper contours and hence detecting the hand sign.



VII. Limitation And Future Research

The study's limitations will apply to any future research on deep learning and sign language recognition that is done in a similar manner. The findings compiled in this paper are sufficient to develop a sign language to text system, however more studies and relevant material should be reviewed for any future research that aims to improve the processes for sign language translation. Although a proposed system was presented in this work, it is not at all practical in real-world situations since it is more of a concept with suggested approaches utilized in isolation. The paper also advises more research into less popular sign language recognition techniques like DNN and CTC.

Although our current system recognizes only ASL, it can be extended to recognize other native sign languages with adequate data sets and training. The system currently implements finger spelling translation, but recognizing contextual signing would require higher processing and natural language processing (NLP). We plan to improve the accuracy of our system in recognizing sign language gestures even in complex backgrounds by experimenting with various background subtraction algorithms. Additionally, we aim to enhance the pre-processing stage to enable better prediction of gestures in low-light conditions. Further development can involve building the system as a web or mobile application for convenient user access.

Furthermore, for creating a better model, implementation can involve not only alphabets or word detection but also complete sentence recognition. Another aspect that could be added to the project would include recognition of different types of sign languages from different countries across the globe, for instance American, Chinese, French etc. sign languages.

IX. REFERENCES:

[1] Sahoo, A. K. (2021, June). Indian sign language recognition using machine learning techniques. In Macromolecular Symposia (Vol. 397, No. 1, p. 2000241).

[2] Thakur, A., Budhathoki, P., Upreti, S., Shrestha, S., & Shakya, S. (2020). Real time sign language recognition and speech generation. Journal of Innovative Image Processing, 2(2), 65-76.

[3] Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-González, A. B., & Corchado, J. M. (2022). Deepsign: Sign Language Detection and Recognition Using Deep Learning. Electronics, 11(11), 1780.

[4] Das, A., Gawde, S., Suratwala, K., & Kalbande, D. (2018, January). Sign language recognition using deep learning on custom processed static gesture images. In 2018 International Conference on Smart City and Emerging Technology (ICSCET) (pp. 1-6). IEEE.

[5] Bankar, S., Kadam, T., Korhale, V., & Kulkarni, M. A. (2022). Real Time Sign Language Recognition Using Deep Learning.

[6] Srivastava, S., Gangwar, A., Mishra, R., & Singh, S. (2021, December). Sign Language Recognition System using TensorFlow Object Detection API. In International Conference on Advanced Network Technologies and Intelligent Computing (pp. 634-646). Springer, Cham.

[7] Sood, E., Tannert, S., Frassinelli, D., Bulling, A., & Vu, N. T. (2020). Interpreting attention models with human visual attention in machine reading comprehension. arXiv preprint arXiv:2010.06396.

[8] Thylashri, S., Femi, D., Jawahar, V., Santhakumar, D., & Reddy, Y. R. (2022, July). Human Sign Language Recognition Using Deep Learning. In 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES) (pp. 1-5). IEEE.

[9] Bharathi, C. U., Ragavi, G., & Karthika, K. (2021, October). Signtalk: Sign Language to Text and Speech Conversion. In 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) (pp. 1-4). IEEE.

[10] Chilukala, M. R., & Vadalia, V. (2022, March). A Report on Translating Sign Language to English Language. In 2022 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 1849-1854). IEEE.

[11] Gupta, V., Jain, M., & Aggarwal, G. (2022, January). Sign Language to Text for Deaf and Dumb. In 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 384-389). IEEE.

[12] Bhat, A., Yadav, V., & Dargan, V. (2022, May). Sign Language to Text Conversion using Deep Learning. In 2022 3rd International Conference for Emerging Technology (INCET) (pp. 1-7). IEEE.

[13] 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.

[14] de Bruijne, M., Cattin, P. C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., & Essert, C. (Eds.). (2021). Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III (Vol. 12903). Springer Nature.

[15] Kulkarni, A. (2021). Dynamic sign language translating system using deep learning and natural language processing. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(10), 129-137.

[16] Patil, R.C., Patil, V., Bahuguna, A., & Datkhile, G. (2021). INDIAN SIGN LANGUAGE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK. Strad Research.

[17] Mr. G. Sekhar Reddy, A. Sahithi , P. Harsha Vardhan , P. Ushasri(2022). Conversion of sign language video to text and speech.

[18] Sreeraksha M R, Vani H Y, Phani Bhushan, D K Shivkumar, 2021, Speech To Sign Language Conversion using Convolutional Neural Networks, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCCDS – 2021 (Volume 09 – Issue 12),

[19] Adithi Krishnan , Ruthvik B R , Spoorthy M , Rhea Muthanna , Shashank N. Sign Language to text conversion.

[20] Kamruzzaman, M. M. (2020). Arabic sign language recognition and generating Arabic speech using convolutional neural network. Wireless Communications and Mobile Computing, 2020.

[21] Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 173, 24-49.

[22] Han, W., Ruan, H., Chen, X., Wang, Z., Li, H., & Schuller, B. (2018). Towards temporal modelling of categorical speech emotion recognition.

[23] Chowdhary, K. (2020). Natural language processing. Fundamentals of artificial intelligence, 603-649.

[24] 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.

[25] Li, G., Hari, S. K. S., Sullivan, M., Tsai, T., Pattabiraman, K., Emer, J., & Keckler, S. W. (2017, November). Understanding error propagation in deep learning neural network (DNN) accelerators and applications. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (pp. 1-12).

[26] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

[27] Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. Journal of Computer science and Technology, 16(6), 582-589.