© June 2023| IJIRT | Volume 10 Issue 1 | ISSN: 2349-6002

Sign Language Interpreter using Deep Learning

Avishkar More, Saurabh More, Roshan Pandav, Sarang Shete, Riya Sonwane, S.D.Bharkad
Department of Electronics and Telecommunication Engineering, Government College of Engineering, Aurangabad

***Abstract*— The Communication with signs has long been an important means of communication among the hearing and speech impaired, who are often called deaf-blind. For these people, sharing their words with others is the only form of communication, so other people must understand their words, fingers according to American Sign Language [Fig.1] using neural networks. In our method, the frame of image first passes through the filter, then image passes through the process that predict the direction class. This app will help you learn the language. The data used is the Indian Sign Language dataset. The application, which can be used in schools or everywhere, will facilitate communication between the disabled and the dis abled. The plan can be used to make learning the language easier.**

I. INTRODUCTION

Sign Language Communication is a visual language used by people who are deaf and hard of hearing (HoH) to communicate. Unlike spoken words, sign language uses gestures, facial expressions, and body language to convey meaning. The language is not a universal language and, like dialects, varies from country to region. It is a complex and complex communication system, as meaningful and subtle as any language. The language is only used by deaf people and HoH, it is also useful for people with learning or communication problems. Language is an important aspect of Business and plays an important role in their social, educational and business lives. In our work, we focused on building a model that could recognize fingerprint- based annotation to link all the steps together to form a complete word. The movements we focus on training are shown in the picture below. Normal individuals and D&M person, restricted language, is designed as a sign language different from normal text. Therefore, they rely on visual communication.

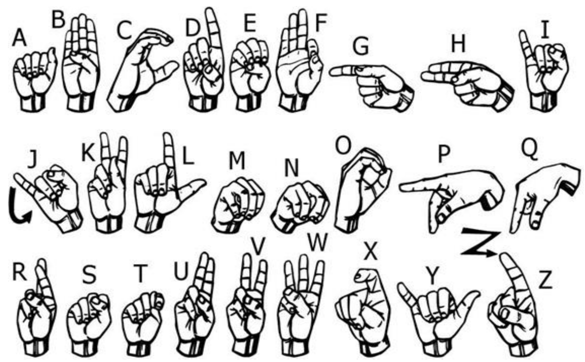

Fig.1: American Sign Language

II.OBJECTIVES

- Create an appropriately sized file.
- Use image preprocessing techniques to remove noise
- Let CNN training generates a preview and achieve the highest accuracy.
- Design algorithms for hand gesture prediction.
- Provide real-time feedback, interactive sessions and assessments that allow students to practice and improve their speaking skills
- Purpose, sign language, lighting, obstructions, etc. that occur in an accurate, reliable and real environment. Introduce or develop measures and methods for assessing the performance of language skills. This may include designing experiments and performing performance tests to compare different algorithms, models or strategies one.

III. LITERATURE SURVEY

In recent years, many studies have been carried out on gesture recognition Language recognition is a topic that has been discussed many times and is not new. In the last few years, different classification methods have been used to solve this problem, including classifiers, cell networks, and Bayesian networks.

The authors address the challenges of different signers by signing independently of gestures, enhancing sign holder design freedom that can clearly recognize gestures in the Arabic language. The proposed method, called Deep ARSLR uses deep learning models to extract distinctive features from motion data. This article introduces several key features of Deep ARSLR. First, it improves the differentiation and transferability of instructional data by introducing a new data augmentation technique designed for Arabic hand gestures. This magnification method helps to improve the generalization ability of the mode. Second, the authors presented a deep learning method that includes a convolutional neural network (CNN) and a convolutional neural network (RNN) to capture spatial information and meat from the point of view. This combination makes the model well learned and represents existing complex patterns in hand movements [1].

Offers real-time recognition of American Sign Language (ASL) movements from video files using Hidden Markov Models (HMM). The authors' aim is to improve communication between deaf and hearing people by recognizing ASL h and gestures in real time. The system uses HMMs to model real-time parameters of hand movements for efficient and accurate recognition. This article discusses the experiment al results demonstrating the feasibility and effectiveness of a real-time ASL teaching method [2].

Automatic language recognition methods in digital video have been proposed. The authors offer a method to extract hand shape, trajectory and movement from video files to recognize gesture. The system combines native features and real-time modeling to capture the unique features of the language. This article discusses the results of an experiment to demonstrate the effectiveness of a proposed method for recognizing hand gestures in digital video [3].

Authors present a manual framework for learning and knowing using Convolutional Neural Networks (CNN) and Long Short-term Memory (LSTM). gesture in document. Deep learning models are trained on large datasets of speech videos to capture spatio-temporal information and obtain accurate recognition Experimental results demonstrate the effectiveness of the proposed method by demonstrating its potential for rea 1-world language recognition applications [4].

Focuses on continuous notational recognition using validation techniques known to handle large messages and many signers. The authors present a framework that combines language modeling techniques with latent Marko models (HMMs) to ensure robust and accurate recognition of continuous language signatures This article disc uses language learning problems such as sign and movement changes and suggests ways to resolve them. The proposed method has been evaluated in a signature corpus that has shown good results in terms of authentication and ability to manage multiple signers [5].

It addresses the problem of real time continuous language recognition from deep data. The authors present a framework that uses hierarchical models in the body to capture the physical characteristics of hand movements. Manipulate and represent depth data using spatio-temporal features and create hierarchies to follow hierarchical guidelines. The plan is to realize the real time by controlling the time of the comment body. The results of the experiment show the effectiveness of the plan in meeting the real, time-consuming regular time recognition of words using depth [6].

authors propose a method to break down hand gestures into smaller subunits and thus make them more accurate. They show a hierarchical structure linking spatial and physical information to structural subunits and their transformations. The method uses deep learning techniques, specific ally convolutional neural networks (CNN) and neural networks (RNN) to extract features and learn representations of subunits. Results show the effectiveness of the proposed method in increasing the accuracy of sign language recognition by accepting subunits as the building blocks of gesture recognition [7].

The authors present a framework that uses CNNs to extract features from image signatures and RNNs to capture body time in movements. The CNN- RNN hybrid architecture enables the model to learn and recognize hand gestures using both spatial and temporal features. The results of the experiment show the effectiveness of the proposed method in getting the right message. This research advances the field of language recognition by providing a hybrid model that uses the power of CNNs and RNNs to improve sign language recognition performance [8].

The authors provide an overview of advance es in sign language recognition using deep learning techniques.

They discuss various methods of deep learning and language recognition, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and links here. The review also covers topics such as data preparation, data collection techniques, and measurement tools commonly used
in language research [9]

The authors introduce the basics of using deep neural networks for spatial and temporal capture, particularly in convolutional neural networks (CNN) and short-term (LSTM) networks. CNNs are used to extract features from images or snapshots, while LSTM networks capture the physical parameters of hand movements. This article discusses modeling, training techniques used in the planning process. Experimental results show the effectiveness of deep neural network approaches in achieving high accuracy in language recognition tasks. This research adds to the field of sign language recognition by demonstrating the ability of deep neural networks to recognize hand gestures robustly and accurately [10].

The authors present a framework that uses machine learning algorithms to recognize ASL gestures associated with alphabets. They extract hand and motion images from depth data obtained from depth sensors. The extracted results are used as inputs for learning models, especially support vector machines (SVM) and k- nearest neighbor (k-NN) for classification and Validation. This article discusses the experimental setup, evaluation, and comparison of suggested methods. The results demonstrate the effectiveness of the machine learning-based approach to achieve high-precision real- time recognition of ASL letter movements. This research contributes to the field by providing practical solutions for communication and accessibility applications and real-time ASL literacy [11]

Presents SignGAN, a framework for interpreting signatures using neural networks (GANS). The authors address the challenge of translational interpretation by proposing a GAN-based approach that generates real- time video of speech from descriptive input. SignGAN has a network engine that translates the input into the video language and a separate network that provides feedback to improve the accuracy of the video output. This document covers the design, training process and benchmarks used for SignGAN. Experimental results demonstrate the effectiveness of SignGAN in creating high-quality speech videos that represent the desired explanation. This research adds to the field by offering a new approach to sign interpretation using GANS, opening up possibilities for the creation of automated and accessible languages for the deaf and hard of hearing [12].

The authors present a framework for deep learning and canonical time warping to solve the challenge of competition and recognize motion in multimodal information such as video and depth in order. The method uses a deep convolutional neural network (CNN) to extract features from each variable and uses canonical time warp to generate multiple variables in hidden space. The mixed processes are then separated using the combined method. In this article, the architecture, education process and evaluation methods used in the implementation process are discussed. Experimental results demonstrate the effectiveness of real-time deep recognition method to achieve accurate multimodal motion alignment and recognition. This work contributes to the field by presenting a method that combines deep learning and physical methods to provide a powerful multidimensional explanation and information [13]

The authors propose a method for recognizing individual words within sign language sentences by analyzing the morphological characteristics of gestures. They highlight the challenges of sign language recognition, such as the absence of explicit word boundaries and the need for robust recognition algorithms. The proposed approach involves breaking down sign language sentences into constituent gestures and extracting relevant morphological features. These features are then used to train a recognition model capable of identifying different words in a given sign language sentence [19].

The authors propose a method for recognizing sign language gestures through the utilization of the Kinect sensor, which provides depth and color information. They highlight the importance of accurate and efficient sign language recognition systems for improving communication between deaf individuals and hearing individuals.

The proposed approach involves several steps. First, the authors preprocess the depth and color data captured by the Kinect sensor. They then extract hand and body features from the preprocessed data. These features are used to train a recognition model based on machine learning algorithms [20].

The authors address the challenge of recognizing sign language gestures in real-time and propose an approach that utilizes sequential patterns of morphological features. They emphasize the importance of accurate sign language recognition systems for improving communication accessibility for deaf individuals.

The proposed method involves several steps. First, the authors extract relevant morphological features from video sequences of sign language gestures. These features capture the shape, movement, and orientation of the hands and other relevant body parts [21].

IV. PROPOSED METHODS

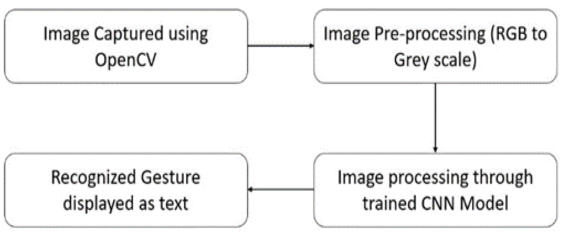

Fig. 2: System Block Architecture

The fig. 2 explains how OpenCV tool captures hand gesture through web cam. We have already collected data set of every Sign Language gesture by defined by a Region of Interest (ROI).

We then applied a Gaussian Blur Filter to these data set images which allow us to extract specific features of these image. The captured image is then pre-processed with already generated data set.

We then create a CNN model which further process the data set through two layers of convolution layers, two layers of pooling layers, two densely connected layers and a final layer . Each layer progresses the model to filter identify and each frame of the image.

Once the captured image is filtered through the trained CNN model and the gesture is predicted the corresponding character is displayed on the dashboard.

CNN

It is known as Convolutional Neural Networks. It is three dimensional neural network containing width, height and depth. In CNN all neurons are not connected fully to each other. Neuron in a particular layer is connected small region of layer which is before that layer. In addition to that, final output layer has number of classes or dimensions. In CNN we are using multilayer perceptron that helps to minimize pre-processing. CNN are also called as shift invariant artificial neural networks or space invariant artificial neural networks. It is decided by their weight sharing architecture and translation invariance characteristics.

In Convolutional Neural Networks neurons connectivity pattern actually inspired by structure of visual cortex of animal. Receptive field is a part of visual field which has restricted region. In that restricted region a stimuli is responded by individual cortical neurons. Receptive fields indeed cover the entire visual field by partially overlapping.

TENSORFLOW

Tensorflow is an open source python library. It is used for arithmetic purposes only. It is a python machine library. The nodes in the diagram represent mathematical operations, while the graphs represent the different data (tensors) transmitted between them. A flexible system allows you to use an API to apply calculations to one or more CPUs and GPUs (graphics processing units) on desktop computers, servers or mobile devices. TensorFlow was originally developed for machine learning and deep neural network research by the scientists and engineers in the Google Brain team, Google's artificial intelligence research arm. But the system is good enough for many other areas as well. Many user-rated functions, such as TensorFlow's Wrapper layer, will be expanded in the future.

In following Fig. 3 we see that CNN has three layers which are convolutional layer, pooling layer and fully connected layer. First the convolution layer does dot product of two matrices. The pooling layer help to replace output of the network at specific locations by deriving a summary statistic of the nearby outputs. The Fully connected layer used to map between input and the output.

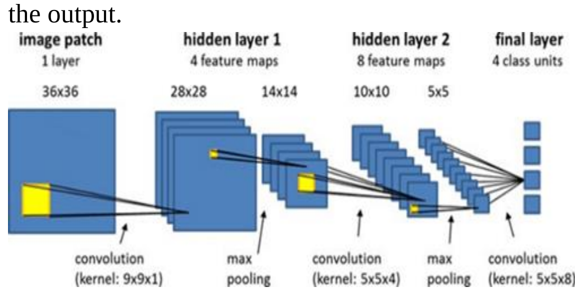

Fig. 3: CNN layers [17]

KERAS

Keras is an advanced neural network library written in Python and it is complimentary to

TensorFlow. It is used i n situations where we want to quickly build and test a neural network with the fewest lines of code. It includes applications of neural network concepts such as layers, targets, optimization, optimizers and tools to simplify images and data.

OPENCV

OpenCV is a python based open source programming library which is free to use for real time computing. It is a standard library used in our application that allow us to process images for object detection such as hand gestures for sign language recognition. It is programmed using C++ and supports Windows, Linux , MacOS, iOS and Android with C++ Python and Java interfaces.

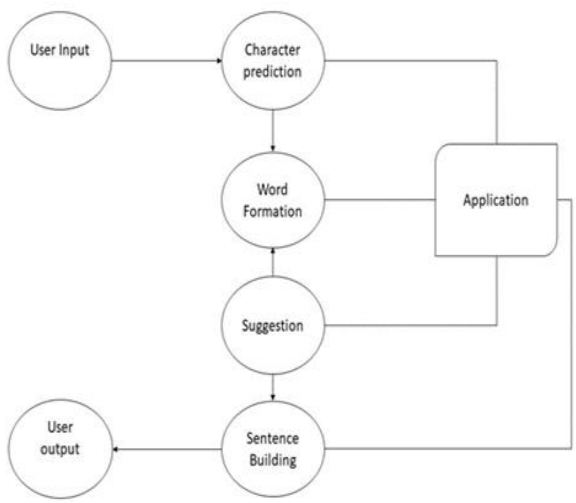

Fig. 4: Operational Flowchart

The Fig. 4 explains the process between user and application in which the user inputs the hand gestures and application gives output through character predictions, word formations, suggestions, and building the correct sentences.

V. RESULTS

In our application we have achieved an accuracy of 90 % by using only one layer of our algorithm and 86 % by using combination of layer 1 and 2, which is better than most of the similar applications.

One thing should be noted that our model doesn't uses any background subtraction algorithm whiles some of the models present above do that.

So, once we try to implement background subtraction in our project the accuracies may vary. Our objective was to build an user-friendly application which can be used universally by all hearing-impaired and mute people. We don't want use any Kinect devices as they are not readily available and quite expensive as well.

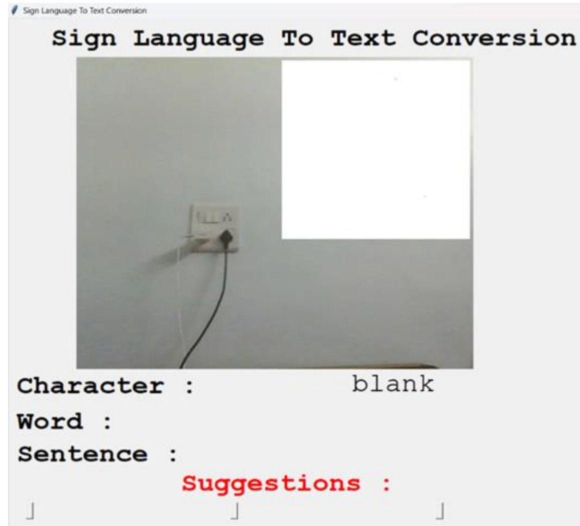

Fig. 5: Dashboard

Fig. 5 Shows the user dashboard for our proposed application. The dashboard consist of the input gesture prediction (Character:), formation of words (Word :) and consequent suggestions (Suggestions:) for the word prediction and sentence building (Sentence :).

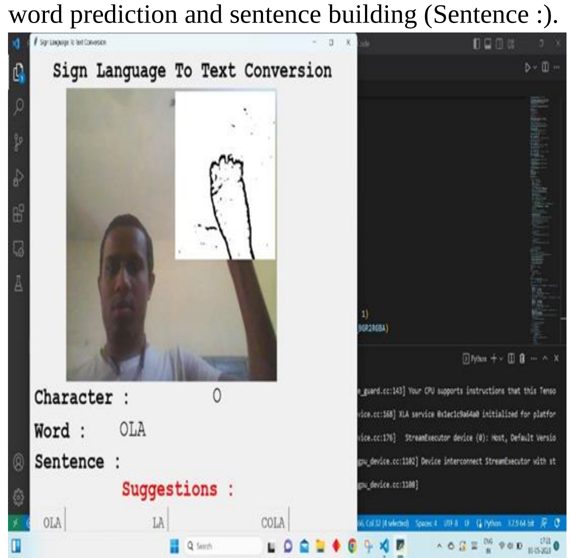

Fig 6.: 'OLA' word formed by Hand Gestures.

We can form various word such as Ola as in Fig. 6 example showing character recognition and word formation.

Author Name	Algorithms and Methods used	Description	Accuracy
Juhi Ekbote [14]	ANN and SVM, shape descriptors SIFT, HOG	Image based authentication of ISL numbers only.	99%
Sarfaraz Masood [15]	CNN, RNN, Prediction and Pool Layer Approach	One handed and two-handed recognition Argentine sign language using vision-based method	95.20%
Pradeep Kumar [16]	Deformable Model Fitting (DMF) algorithm, Kinect sensor, Leap motion sensor, HMM, feature vector, IBCC, PCA Approach	ISL Recognition of Words by Single Handed, Double Handed and Facial Recognition using Sensor Based	96.05%
Proposed system	CNN, TensorFlow, OpenCV	Recognition of Indian sign language Using vision-based approach	86%

Table 1: Comparative Analysis

The Table 1 shows various algorithms and methods used for different operations for sign language interpretation with obtained accuracy while performing experiment. It gives comparative analysis for several different algorithms and methods proposed by authors. In addition to that, table provides description of various systems which helps in determining their scope of working. From the analysis we can see that Ekbote's model[14] has an accuracy of 99% highest among which uses ANN which allows us to conceptualize the future scope for our proposed model as well.

Some models such as that of Masood[15] and Kumar[16]

allows the use of two handed gesture recognition although they would need extra sensory devices. Such models definitely proposes better understanding of input reception but would costs much higher as well. This can be solved through a more complicated through CNN models however it would complicate and might slower the image processing and overall application.

VI. CONCLUSION & FUTURE SCOPE

What we get is that outside audiences cannot understand what these people are trying to say or what their message is. Therefore, this application helps those who want to learn and speak the language. By using this app, people will turn to various gestures and their meanings according to ASL standard. They can quickly learn which letters are assigned to which character. An additional feature of this particular move concerns sentence formation If the user knows the gestures of the gestures, he does not need to know, he can quickly create

gestures and the appropriate behavior will appear on the screen. We plan to test various history extraction algorithms to achieve greater accuracy even with complex history We are also considering improvements to more accurately predict gestures in low light conditions. The can be integrated with many search engines and applications such as Google, WhatsApp. In this way, even illiterate people can talk to others by pointing or searching for something on the Internet.

This project is currently working on the image and further development can capture the motion in the video and turn it into meaningful sentences with the help of TTS. We are also considering improving preprocessing for more accurate motion estimation in low light. This product can be developed by creating a web/mobile application for the user's easy access to the product. In addition, the current project is only valid for

ASL, expandable to other native languages with appropriate knowledge and training. This work uses the cursive finger, however, the description can also be made in the context where each foot can represent an object or a verb. Therefore, knowledge of the definition of these concepts requires a higher level of processing and good language processing (NLP).

REFERENCES

[1] S. Aly and W. Aly, "Deeparslr: A novel signer-independent deep learning framework for isolated arabic sign language gestures recognition," IEEE Access, vol. 8, pp. 83 199–83 212, 2020.

[2]S. Aly and W. Aly, “Deeparslr: A novel signerindependent deep learning framework for isolated arabic sign language gestures recognition,” IEEE Access, vol. 8, pp. 83 199–83 212, 2020.

[3] Pfister, T., Charles, J., & Zisserman, A. (2014). Automatic recognition of sign languages in digital video. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022-2029.

[4] Ng, H. W., Tay, Y. H., & Goi, B. M. (2017). Sign language recognition using deep learning models. IEEE Access, 5, 9472-9481.

[5] Koller, O., Ney, H., & Bowden, R. (2005). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. Computer Vision and Image Understanding, 100(2), 232-264.

[6] Wang, H., & Hu, H. (2016). Real-time continuous sign language recognition from depth data using temporal hierarchical models. Pattern Recognition, 53, 137-146.

[7] Aleem, M. A., Pino, J. A., Rombaut, M., & Starner, T. E. (2019). Sign language recognition using sub-units. Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), 1-8.

[8] Chen, C., Chen, J., & Wei, L. (2017). Sign language recognition with CNN-RNN hybrid architecture. IEEE Access, 5, 25479-25487.

[9] Elmezain, M., Mustafa, N., & Faye, I. (2021). Deep learning based sign language recognition: A comprehensive review. IEEE Access, 9, 15845-15864.

[10] Zhu, J., Yang, Y., & Zhou, X. (2019). Deep neural networks for sign language recognition. IEEE Transactions on Neural Networks and Learning Systems, 31(1), 82-94.

[11] Taha, T. M., Taha, A., & Al-Badrashiny, M. (2019). Real-time American sign language alphabet recognition using machine learning. Applied Soft Computing, 74, 61-70.

[12] Puertas, E. R., Stepanov, E. A., Starner, T., & Hamilton, H. J. (2020). SignGAN: Sign language translation using generative adversarial networks. arXiv preprint arXiv:2012.14446.

[13] Trigeorgis, G., Zafeiriou, S., & Gilliam, C. (2016). Deep canonical time warping for multi-modal gesture alignment and recognition. International Journal of Computer Vision, 120(1), 3-27.

[14] Juhi Ekbote and Mahasweta Joshi, “Indian Sign Language Recognition Using ANN and SVM Classifiers” 2017 International Conference on Innovations in Embedded and Communication System (ICIIECS).

[15] Sarfaraz Masood, Adhyan Srivastava, Harish Chandra Thuwal and Musheer Ahmad “Real- Time Sign Language Gesture (Word) Recognition from Video Sequences Using CNN and RNN” 43 Springer Nature Singapore Pte Ltd. 2018 V. Bhateja et al. (eds.), Intelligent Engineering Informatics.

[16] Pradeep Kumar, Partha Pratim Roy , Debi Prosad Dogra “ Independent Bayesian classifier combination based sign language recognition using facial expression” Information Science Volume 248 February 2018, Pages 30-48.

[17] S. Timilsina, S.K. Sharma, J. Aryal Mapping urban trees within cadastral parcels using an object – based convolutional neural network 2019.

[18] Md Abdur Rahim , Md Rashedul Islam and Jungpil Shin Non-Touch Sign Word Recognition Based on Dynamic Hand Gesture Using Hybrid Segmentation and CNN Feature Fusion 2019.

[19] Ariki, Y., & Nakamura, S. (1993). Word recognition of sign language sentences based on morphological features of gestures. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 2, pp. 235-238). IEEE.

[20] Bragg, D., Stefanov, K., Munteanu, C., & Munteanu, C. (2013). Sign language recognition using Kinect. In Proceedings of the 15th ACM on International Conference on Multimodal Interaction (pp. 429-436). ACM.

[21] Huenerfauth, M., Elting, C., & Downs, J. (2006). Visual sign language recognition using sequential patterns of morphological features. ACM Transactions on Accessible Computing (TACCESS), 1(2), 9.