Sign Language Interprter

Mrs. Akshatha AMS, Abhishek TM, Sourav HS ,Suhas BM, Sri Krishna Dev Rayudu

Department of Artificial Intelligence and Machine Learning, G M Institute Of Technology, Davanagere, Karnataka 577 004

Abstract

Sign language is a form of communication commonly used by people with hearing impairment or people with speech impediments. Not all ordinary people understand the language. The translation of sign language into the alphabet/text automatically will facilitate the communication of the deaf with ordinary people. In recent years, deep convolutional networks (DCNs) have shown promising results in sign language recognition, owing to their ability to learn and extract hierarchical features from the input data. This report explores the application of DCNs in sign language recognition and presents an overview of various DCN architectures used in this domain. The report also highlights the challenges and limitations of DCNs in sign language recognition and discusses future research directions in this field. Finally, the report provides a detailed evaluation of the performance of DCNs on different sign language datasets, which demonstrates their effectiveness in accurately recognizing sign language gestures

Keywords: Sign language, Convolutional network, DCN, datasets

1. INTRODUCTION

The world can't exist without correspondence, whether or not it appears as contact, discourse, or visual articulation. Text and visual articulations work with correspondence between the hard of hearing and the quiet. Hands and facial highlights are exceptionally critical in offering human viewpoints in confidential correspondence. Various mechanical upgrades and much examination have been directed to help not-too-sharp people. Profound learning and PC vision can likewise be used to affect this reason.

If an individual can't talk or hear, gesture-based communication is the main method for correspondence accessible to them. Fingerspelling is a vital tool in sign language, as it enables the communication of names, addresses, and other words that do not carry meaning in word level association.

Gesture-based communication permits provoked people to offer their viewpoints and sentiments. In this paper, a special gesture-based communication recognizable proof strategy for distinguishing letter sets and movements in communication through signing is proposed.

The problem we are investigating is sign language recognition through unsupervised feature learning. Many systems are developed for sign language recognition and there is no exact system for recognizing the complete signs. Being able to recognize sign language is an interesting computer vision problem while simultaneously being extremely useful for deaf people to interact with people who don’t know how to understand American Sign Language (ASL).

2.LITERATURE SURVEY

Translation of Sign Language into Text Using Kinect for Windows, P. Amatya, k.Sergieva & G. Meixener, This model proposes methods to recognize and translate dynamic gestures of the German Sign Language (Deutsche Gebärdensprache, DGS) into text using Microsoft Kinect for Windows v2. Two approaches were used for the gesture recognition process: sequence matching using the Dynamic Time Warping algorithm and a combination of Visual Gesture Builder along with Dynamic Time Warping. For benchmarking purposes, eleven DGS gestures, which were provided by an expert user from Germany, were taken as a sample data set. The proposed methods were compared based on the computation cost and accuracy of these gestures. The computation time for Dynamic Time Warping increased steadily with an increasing number of gestures in the data set whereas in the case of Visual Gesture Builder with Dynamic Time Warping, the computation time remained almost constant. However, the accuracy of Visual Gesture Builder with Dynamic Time Warping was only 20.42% whereas the accuracy of Dynamic Time Warping was 65.45%. Based on the results, we recommend the Dynamic Time Warping algorithm for small data sets and Visual Gesture Builder with Dynamic Time Warping for large data.

American sign language translation using edge detection and cross correlation, A. Joshi, H. Sierra & E. Arzuaga, American Sign Language Translation Using Edge Detection and Cross Co-relation This project is to implement an automated translation system that is capable of 12 translating ASL to English text using common computing environments such as a computer and a generic webcam. In this project, a real-time hand gesture recognition system using a combination of image processing modalities is implemented. A prototype graphical user interface application for ASL sign capture, processing, collection, and analysis is presented. The approach consists of a gesture extraction phase followed by a gesture recognition phase. An image gesture database is collected through the application and used as training information to be used in the gesture recognition stage. This model aims to provide two different translation paradigms:

◦ English Characters (alphabet)

◦ Complete words or phrases

In the method to recognize individual characters, the hand gesture image is processed by combining image segmentation and edge detection to extract morphological information and then processed by the gesture detection stage that recognizes the corresponding alphabet letter. In this feature selection stage, a subset of frames that can represent a particular word or phrase is selected. The collection of frames representing a word or a phrase is then processed using the multi-modality technique used for processing individual characters. Finally, the gesture recognition stage is applied to both approaches using a cross-correlation coefficient-based scheme to detect the expression

Image processing-based method for extraction of descriptors followed by a hand shape classification using ProbSom which is a supervised adaptation of self organizing maps, Ronchetti et al. The study titled "Image processing-based method for extraction of descriptors followed by a hand shape classification using ProbSom" by Ronchetti et al. presents an innovative approach to sign language recognition. The authors utilized image processing techniques to extract descriptors representing key features of hand shapes. These descriptors were then inputted into ProbSom, a supervised adaptation of self-organizing maps, for classification Remarkably, the technique yielded impressive results in two distinct sign languages. In Argentinean Sign Language. (ASL), the accuracy of recognition exceeded 90%, showcasing the effectiveness of the proposed methodology. The classification process in ASL was based on the eigenvalue-weighted Euclidean distance, a robust metric for comparing and distinguishing between hand shapes. Additionally, the study extended its application to Indian Sign Language (ISL), successfully identifying 24 different alphabets with an accuracy of 96%. This highlights the versatility and reliability of the proposed approach across different sign languages.

3.PROPOSED SYSTEM

1)**Dataset Collection:** Collect images of hand gestures representing ASL alphabets using a camera integrated with the Mediapipe hand tracking library.

2)**Model Training:** Utilize a Convolutional Neural Network (CNN) architecture to learn and recognize ASL alphabet gestures from preprocessed images**.**

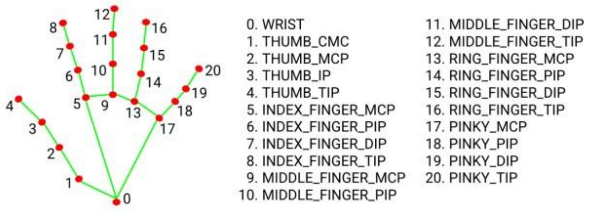
3)**Testing and Verification**: Evaluate the trained CNN model on a separate test dataset to assess performance in recognizing ASL gestures accurately.

4)**Text-to-Speech Integration:** Implement a text-to-speech synthesis feature to convert recognized ASL gestures into spoken words for effective communication.

5)**Words Builder:** Develop a words builder feature to construct complete words from individual letter gestures, enhancing communication usability.

4.METHODOLGY

Dataset Collection The dataset for American Sign Language (ASL) recognition was obtained by capturing images of hand gestures representing every alphabet using the Mediapipe hand tracking library. The hand-tracking library provided a fast and reliable way to detect hand landmarks and track hand movements in real time. The captured hand gestures were then mapped onto a white background image, which provided a consistent and neutral background for image processing and classification. The resulting dataset can be used to train machine learning models for ASL recognition and assistive technology applications. Mediapipe hand tracking library uses a deep learning-based approach to detect and track human hands in real time. The library is based on a lightweight convolutional neural network (CNN) model, which is trained on millions of annotated hand images to learn the hand landmarks and their connections. The network is optimized for efficiency and can run on mobile devices and desktop computers. The library first detects the hand regions in the input image using a bounding box regression algorithm. It then feeds the detected hand regions to the hand landmark model to estimate the landmarks of each hand. The hand landmarks are a set of 21 2D points that represent the joints and fingertips of the hand. The hand landmark model is a feed-forward neural network, which takes the detected hand region as input and produces the landmarks as output. The model is trained using a combination of synthetic and real data to generalize to different hand shapes, skin colors, and lighting conditions. Once the hand landmarks are detected, the library can use them for various applications, such as gesture recognition, hand pose estimation, and augmented reality. The library also provides a set of utilities for visualizing and processing the hand landmarks, such as drawing hand annotations, calculating hand features, and filtering noisy landmarks.

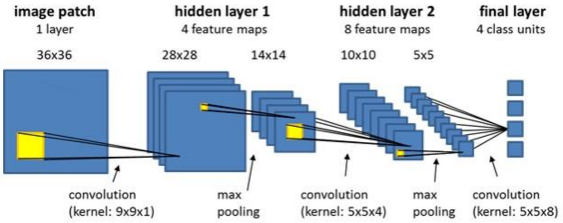


Mediapipe's landmark system

Algorithm Used:

Convolutional Neural Network (CNN): A Convolution Neural Network (ConvNet/CNN) is a Deep Learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlaps to cover the entire visual area. Convolutional neural networks or ConvNets are great at capturing local spatial patterns in the data. They are great at finding patterns and then using those to classify images. ConvNets explicitly assume that input to the network will be an image. CNNs, due to the presence of pooling layers, are insensitive to the rotation or translation of two similar images; i.e., animage and its rotated image will be classified as the same image. Due to the vast advantages of CNN in extracting the spatial features of an image, we have used the Inception-v3 model of the TensorFlow library which is a deep ConvNet to extract spatial features from the frames of video sequences. Inception is a huge image classification model with millions of parameters for images to classify.

Convolutional Layer: In the convolution layer, I have taken a small window size [typically of length 55] that extends to the depth of the input matrix. The layer consists of learnable filters of window size. During every iteration, I slid the window by stride size [typically 1], and compute the dot product of filter entries and input values at a given position. As I continue this process will create a 2-Dimensional activation matrix that gives the response of that matrix at every spatial position. That is, the network will learn filters that activate when they see some type of visual feature such as an edge of some orientation or a blotch of some color.

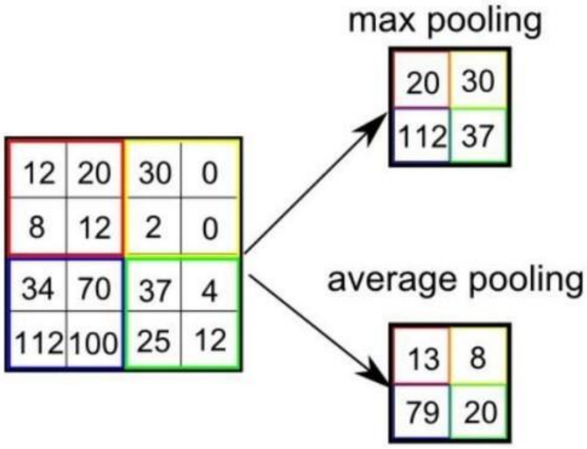


Different convolutional layer

Pooling Layer: We use a pooling layer to decrease the size of the activation matrix and ultimately reduce the learnable parameters. There are two types of pooling:

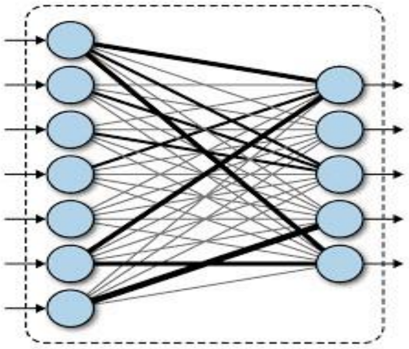
a. Max Pooling: In max pooling, we take a window size [for example window of size 22], and only take the maximum of 4 values. Well, lid this window and continue this process, so we'll finally get an activation matrix half of its original Size.

b. Average Pooling: In average pooling, we take an average of all Values in a window



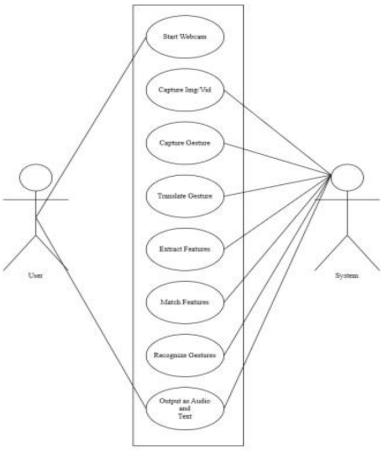
Average pooling and max pooling

Fully Connected Layer: In the convolution layer neurons are connected only to a local region, while in a fully connected region, well connect all the inputs to neurons.

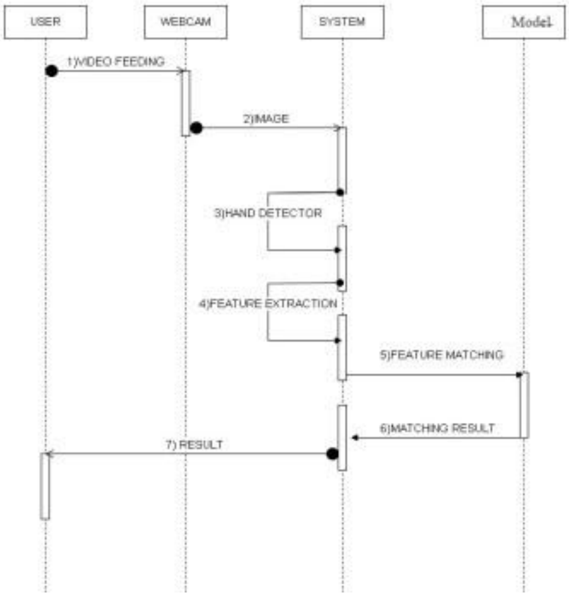


Fully connected layer

c. **Final Output Layer:** In the convolution layer neurons are connected only to a local region, while in a fully connected region, well connect all the inputs to neurons. After getting values from the fully connected layer, well connect them to the final layer of neurons [having a count equal to a total number of classes], which will predict the probability of each image being in different classes.



Use case diagram of Sign language interpreter



Sequence diagram of sign language interpreter

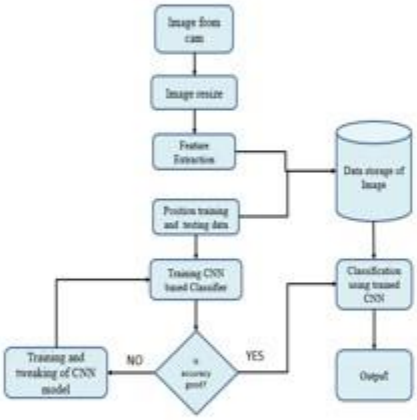
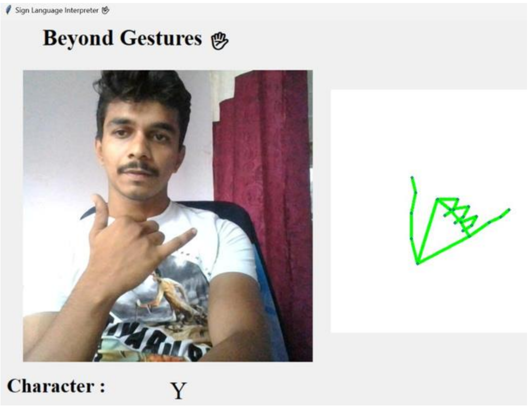


Image from cam
Image resize
Feature Extraction
Position training and testing data
Training CNN based Classifier
Training and tweaking of CNN model
NO
is accuracy good?
YES
Data storage of Image
Classification using trained CNN
Output

Flow chart of Sign Language Interpreter

5. RESULTS

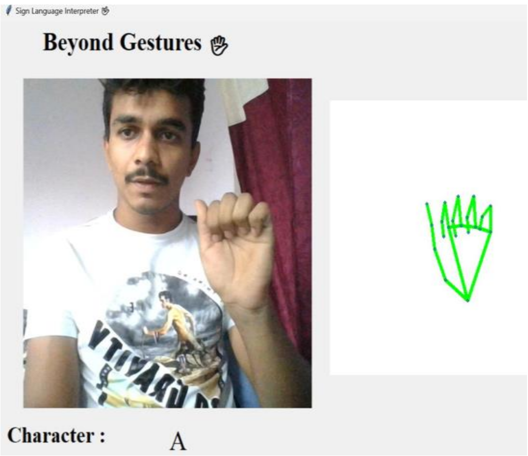


Sign Language Interpreter

Beyond Gestures

Character : Y

Recognition of character Y

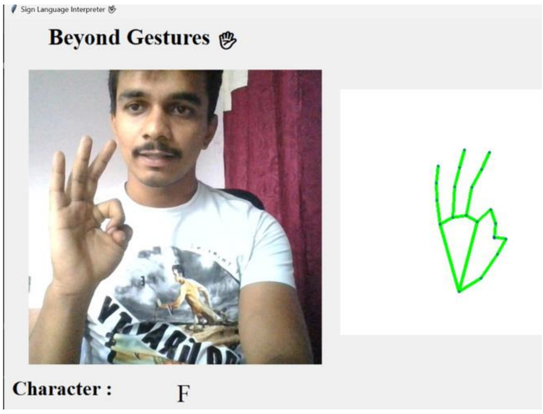


Sign Language Interpreter

Beyond Gestures

Character : A

Recognition of character A



Sign Language Interpreter

Beyond Gestures

Character : F

Recognition of character F

6.CONCLUSION

In this project, we have presented a system for American Sign Language (ASL) recognition using a deep learning-based approach. Our system uses a convolutional neural network (CNN) trained on a dataset of hand gestures mapped to the ASL alphabet. We have demonstrated that our system achieves high accuracy in recognizing ASL gestures in real-time using live video feeds or recorded videos. Our system can be used to aid people with hearing and speech impairments in communicating with others.

7.REFERENCES

[1] Dhiman, M., 2021. Summer Research Fellowship Programme of India's Science Academies 2017. [online] AuthorCafe.
[2] T D, Sajanraj & M V, Beena. (2018). Indian Sign Language Numeral Recognition Using Region of Interest Convolutional Neural Network.
[3] Sharma, R., Bhateja, V., Satapathy, S.C., Gupta, S.: Communication device for differently abled people: a prototype model. In: Proceedings of the International Conference on Data Engineering and Communication Technology, pp. 565–575. Springer, Singapore (2017)
[4] P.Amatya,k.Sergieva & G. Meixener," Translation of Sign Language Into Text Using Kinect for Windows v2". [5] A. Joshi, H. Sierra & E. Arzuaga," American sign language translation using edge detection and cross-correlation".
[6]. Pigou, L., Dieleman, S., Kindermans, P.-J., Schrauwen, B.: Sign language recognition using convolutional neural networks. In: Workshop at the European Conference on Computer Vision 2014, pp. 572–578. Springer Internationa