International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 07 | July 2022 www.irjet.net p-ISSN: 2395-0072

SIGN LANGUAGE INTERFACE SYSTEM FOR HEARING IMPAIRED PEOPLE

Prof. Mrs. Saira Banu Atham[*1], Mr. Chandan H[*2], Mr. Chandan S[*3], Ms. Ashmita Jha[*4], Mr. Devarinti Siva Reddy[*5]

*[*1]Professor, Department of Computer Science & Engineering, Presidency University, Bengaluru, Karnataka, India.*
*[*2,3,4,5]Student, Department of Computer Science & Engineering, Presidency University, Bengaluru, Karnataka, India.*

---***---

Abstract - *Communicating with people is a fundamental form of interaction, and being able to do so with people who have hearing impairments has proven to be a significant challenge. As a result, we are using Sign language as a critical tool for bridging the communication gap between hearing and deaf people. Communicating through sign language is difficult, and people with hearing impairments struggle to share their feelings and opinions with people who do not speak that language. To address this issue, we require a versatile and robust product. We proposed a system that aims to translate hand signals into their corresponding words as accurately as possible using data collected and another research. This can be accomplished through the use of various machine learning algorithms and techniques for classifying data into words. For our proposed system, we are using the Convolutional neural networks (CNN) algorithm. The process entails being able to capture multiple images of various hand gestures using a web cam in this analysis, and the system being able to predict and display the corresponding letters/words of the captured images of hand gestures that have previously been trained with Audio. Once an image is captured, it goes through a series of procedures that include computer vision techniques such as gray-scale conversion, mask operation, and dilation. The main goal and purpose of this project will be to remove the barrier between deaf and dumb people and normal people.*

Key Words: Sign Language, Convolutional Neural Network (CNN), Machine Learning, Hand Gesture, Hearing Impairments.

1.INTRODUCTION

Communication has been and will continue to be one of the most important necessities for basic and social living. People with hearing disabilities communicate each other using sign language, which is difficult for normal people to grasp. The main goal of this challenge is to prevent this by using Sign language to establish communication between deaf and hearing people. As a future scope, this project extension to framing words and commonplace phrases will not only come in handy for deaf and dumb people to communicate with around the world, but it will also aid in the development of independent structures to know-how and helping people.

The majority of countries are developing their own standards and interpretations of various sign gestures. For example, an alphabet from American Sign Language (ASL) will not depict the same factor as an alphabet from Indian Sign Language (ISL). This is a major drawback for many countries. As this emphasizes diversity, it also highlights the complex struggles of sign languages. Deep gesture knowledge should be well-versed in all gestures in order to achieve reasonable accuracy.

2. EXISTING SYSTEM

In [1] they have proposed a system for sign language recognition using CNN algorithm. It involves a series of steps where after preprocessing the image, the trained CNN model compares it to previously captured images and displays the predicted text for the image.

And in [2], they have used a previously trained user defined dataset. They trained not just alphabets but also few specific words which is helpful in real-time experience with the help of CNN model. For recognizing words, they have used ConvNets to attain maximum accuracy which handle 2D or 3D data as input. This model is also capable of handling large amounts of data.

In [1] they were only able to predict single alphabets, whereas in the proposed system in [2] had moved a step forward and was able to predict words with advancement.

In order to attain maximum accuracy [3] have introduced filtering of images. This improves the accuracy of identifying symbols in different low light areas. Here before the image begins process of saturation and grey scaling, filtering takes place first to find the symbol shown in the hands.

Swapna and S Jaya [4] have performed fundamental research on different types of sign language recognition happening like using Image based sign language translator, Glove based sign language translator and sensor gloves. After taking into consideration all types, they have used 5DT (Fifth Dimension Technologies) gloves and as their proposed system. With the help of KNN, Decision tree classifier and neural network they have been able to achieve the required accuracy and recognition.

[5] have proposed a system in which they have used HSV. Once the image is captured using webcam, it passes through HSV to detect and eliminate the background. After which segmentation is done in the region of skin tone. OpenCV is used to remove the difference between different captured

images and amended to same size. It further undergoes dilation and erosion using elliptical kernel. CNN is applied for training and classification after the binary pixels are extracted from the frames.

3. PROPOSED SYSTEM

Our proposed system follows different steps, in which it takes the image from video feed and provides the output as text and audio.

In the first phase, it captures the image from a video feed and provides it as input; subsequently, various image processing techniques are used to eliminate noise and background from the image. To avoid errors while training the CNN model, all collected pictures are resized to the same dimension. We use the RGB-HSV conversion method to remove noise and background from the image, which helps to improve image quality by adding brightness to the image. For hand gesture recognition, we use the CNN algorithm, which provides more accurate results than the KNN and decision tree algorithms. We also use a real-time sign language recognition system and a CNN 2D model, which enhances the model's accuracy. We have added additional images to our dataset to improve prediction. Our method does not include any hand glove or sensor-based approaches, as they might make it difficult for people to use the device. We have also added an extra aspect to it, which is that the obtained output is in the form of text/letter, which we are converting into speech for good understanding and user interface.

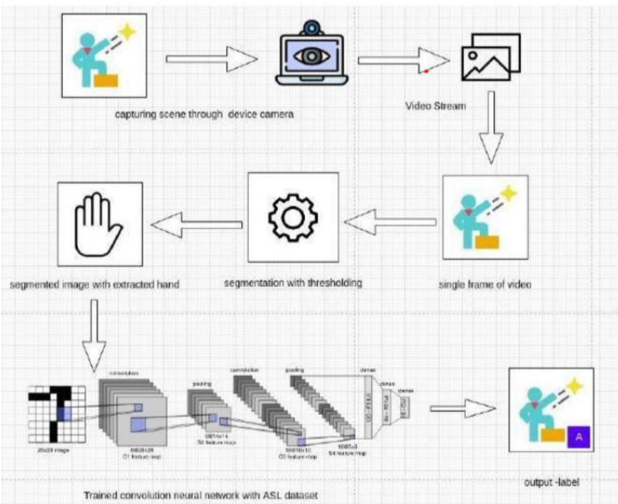


Fig -1: System Flow Chart

4. IMPLEMENTATION

The Sign language recognition system that we are proposing, detects diverse hand motions by recording video and turning it into frames. The pixels are then split, and the picture is generated and submitted to the learning algorithm for assessment. As a result, our method is more resilient and precise in obtaining correct captions for letters. Convolutional neural networks (CNN) algorithm is used for image classification. HSV(Hue, Saturation, Value) and grey scale conversion techniques used for image preprocessing. The **Fig-2** represent hand symbols/gestures used in our system.

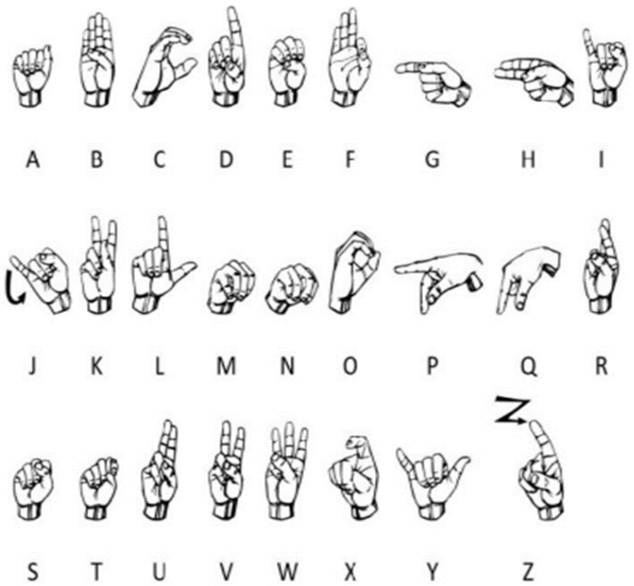


Fig -2: The American Sign Language Symbols

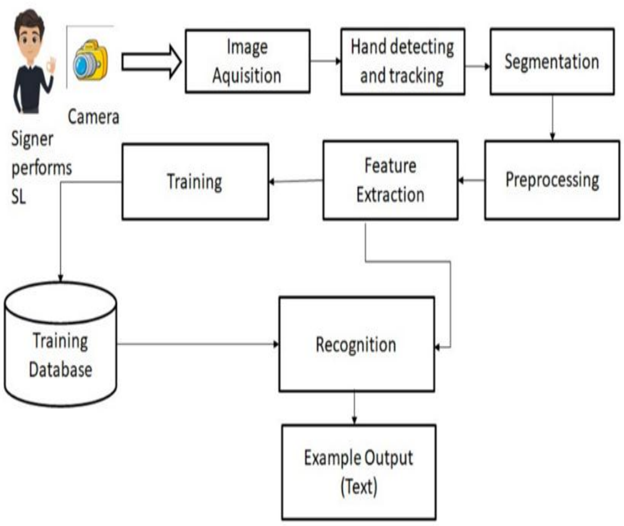


Fig -3: System Workflow

4.1. IMAGE ACQUISITION

It is the process of getting an image from a source, usually a hardware-based source, for image processing. In our project, the hardware-based source is Web Camera. Since no processing can be done without an image, obtaining it is the initial stage in the workflow sequence. The image acquired has not been modified in any way; it is simply a raw image, we use OpenCV python library for capturing image/video from webcam. Here we capture 1200 sample images and store it in our database.

Fig -4: Captured Input Image

4.2. SEGMENTATION & FEATURES EXTRACTION

This approach is concerned with removing items or signs from the context of a taken image. It includes a number of techniques, such as skin-color detection. Edge detection and context subtraction. The motion and position of the collected image must be identified and split in order to identify signals. Form, contour, geometrical feature (position, angle, distance, etc.), color feature, histogram, and other specified properties are taken from input images and applied afterwards for sign classification or identification. A phase in the image compression technique that splits and classifies a big collection of original data into smaller, simpler classes is feature extraction. Hence , processing would be more straightforward. Most crucial element of these big data sets is the wide range of variables. A substantial amount of computer power is required to handle these variables. As a consequence, function extraction assists in selecting the best feature from large sets of data by choosing and merging parameters into functions and lowering data size. These features are simple to use correctly and uniquely while defining the data collected.

Fig -5: Image in ROI

4.3. PRE-PROCESSING

Each image frame is refined to remove noise using a range of filters such as erosion, dilation, and Gaussian smoothing. When a color image is converted to grayscale, its size is decreased. Converting an image to gray - scale images is a valuable way for decreasing the quantity of data to be processed. The next step is to convert grey-scale images to HSV. The images are converted to grayscale and then to HSV before being fed into the skin masking tool. HSV may be used to define an image's color space (Hue, Saturation, and Value). The hue in HSV specifies the color. The color in this model is an angle ranging from 0 to 360 degrees. The range of grey in the color space is determined by saturation. Value denotes the brightness of a color that fluctuates with color saturation. This process removes the extra backdrop and allows you to focus on the hand region that needs to be used. We set the upper and bottom bounds for the picture based on the HSV output.

Fig -6: RGB to HSV Image Conversion

4.4. CNN WORKFLOW

CNNs are a basic form of deep learning, where a more delicate model advances the progress of artificial intelligence by providing systems that imitate many aspects of biological human neural activity. The CNN model is fed with the processed pictures to classify the images. We are obliged to use the American Sign Language alphabet dataset to perform the image processing techniques. Convolution Neural Network was used to train the pictures. We tend to feed the CNN model with the processed images to classify the images. Once the model is fitted, we tend to use it to predict random images from the computer to predict the text and show the output on the screen.

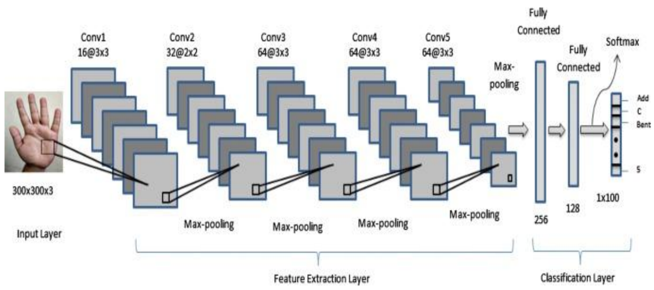


Fig -7: CNN Architecture

CNN has different layers through which the image goes for further classification. Initially we are using CNN model along with sequential classifier.

• Convolution Layer 1 :

In the convolutional layer, it is converted into 2D model of 32 bits 3x3 and the input image of 64x64 pixels of 3 bits with “relu” activation function. The image is reduced into a smaller matrix to reduce the difficulty in computing large image matrix which are obtained from convolution operation. The complexity is reduced by down sampling it.

• Pooling Layer :

Pooling layer is used to reduce the dimensions of the feature maps. Here the maximum pixel value is selected. Here pool size for 'MaxPooling' function for the 2D model is set to 2x2.

• Convolution Layer 2 & 3:

After this a second convolutional layer is introduced with a classifier of 2D model with 32 pixels of 3x3, using 'relu' activation function. Along with a max pooling layer with a pool size of 2x2 as the first convolutional layer. A third convolutional layer is introduced with 64 pixels of 3x3 and the same pool size of 2x2 as previous layers.

• Flattening :

The next step is Flattening where we use 'classifier. Add(Flatten())’ to complete the process. . This method is transforming a pooled function map into a single column which can be easily passed to the final layer that is the Fully Connected layer. Here each neuron corresponds to pixels and these pixels are chosen randomly. It is linked through its backpropagation process.

• Fully Connected Layer :

At this stage, as a name suggest fully connected: every neuron in the network is connected to each otherwhere. Here the classifier with dense layer, 'relu' and 'softmax' activation function is used.

• Compiling:

After going through all the layers, the CNN model is compiled for errors and optimization. We use different metrics to find the accuracy of the model. Lastly, we save the model to our directory for testing the captured images.

4.5. OUTPUT

After getting the output text, it is converted into speech for better understanding

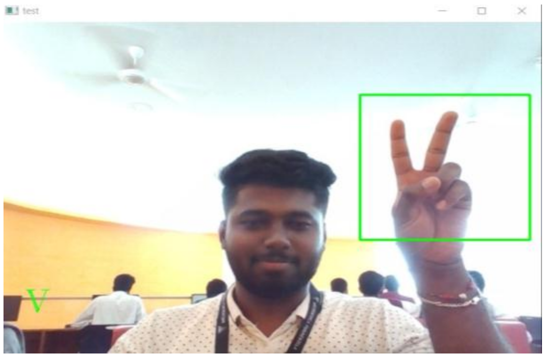


Fig -8: Output Screen

5. EVALUATION

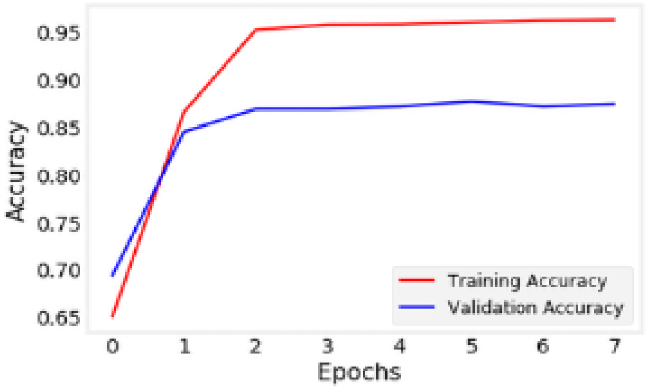
Below **Table -1** shows the input images and its corresponding output images with text. We tested the model for all the alphabetical hand gestures of American sign language and obtained around 95% accuracy. Here we have showed output for letters A,B,C,G and V. First and second column shows actual alphabet and input image, while third and fourth column shows RGB-HSV Converted image and predicted alphabet. At last, the predicted text is converted into audio/speech by using python package.

Alphabet	Input Image	Output Image	Output Obtained
A			A
B			B
C			C
G			G
V			K

Table -1: Output

After collecting the output from the CNN model, we build the accuracy graph and analyses it to gain a better knowledge of the dataset's training and testing.

The accuracy graph **Graph -1** below displays the accuracy of each epoch. The graph is created by processing the image using the approach described in Section 4.3. The training accuracy at the last epoch is 95%, while the testing accuracy is 85%.



Graph -1: CNN Accuracy graph

6. LIMITATIONS

a) While sign language is an excellent mode of communication, it continues to be a challenge for those who do not understand sign language to communicate with mute people.

b) Sign language calls for using arms to make gestures. This shall be a downside for those who do no longer have complete use of their palms or have disabilities.

c) Due to the variations in Sign language fluctuating from one united states of America to every other, it will become difficult to analyze all kinds of gestures.

d) Classification methods are also varying among researchers. Researchers generally tend to develop their own idea, based totally on personal techniques, to give higher bring about spotting the signal language.

e) The model we are using has an ability to do the categorization that is influenced by the model's background and environment, such as the illumination in the room and the pace of the motions and also other factors influence. Due to changes in views, the gesture seems distinct in 2D space.

7. FUTURE SCOPE

As described in the preceding part, our system has various problems to solve. If we can overcome these issues and construct a robust system that can detect low illumination and recognize hand gestures by improving image quality, will have a better scope. Integrating it with some high-end IOT devices will also make installation easier. Sign language does not simply consist of alphabets; it also contains numerals and words, thus adding this element to our system is critical.

8. CONCLUSION

The goal of this project is to bridge the gap between hearing impaired people and regular people. We achieved this by utilizing computer technologies such as deep learning and computer vision. Our system will capture hand gesture images from live video feed and translate them into text/letters that regular humans can understand. This procedure uses a variety of methodologies and technologies; after receiving a picture from the camera, it is converted into greyscale, and just the hand gesture is retrieved from the entire image. The image is then transformed once again into HSV to reduce image quality and eliminate background for better prediction. At last, the processed image is then sent through different layers of CNN algorithm as mentioned in the **section 4.4**, then final text is displayed along with speech.

However, modern technology and methodologies from sectors such as natural language processing and voice assistance are available to explore. Furthermore, although some have imposed grammatical constraints or linguistic limits, others have looked into data-driven solutions, both of which have advantages, given that the linguistics of most sign languages is still under development.

REFERENCES

[1] Teena Varma, Ricketa Baptista, Daksha Chithirai Pandi, Ryland Coutinho, “Sign Language Detection using Image Processing and Deep Learning”, “International Research Journal of Engineering and Technology (IRJET)”, 2020.

[2] Aman Pathak, Avinash Kumar, Priyam Gupta, Gunjan Chugh, “Real time sign language detection”, “International Journal of Computer Science and Network Security”, 2021.

[3] A Sunitha Nandhini, Shiva Roopan D, Shiyaam S, Yogesh S, “Sign Language Recognition using Convolutional Neural Network”, “Journal of Physics”. 2021.

[4] Swapna Johnnya , S Jaya Nirmala, “Sign Language Translator Using Machine Learning”, “ResearchGate”, 2022.

[5] Mehreen Hurroo, Mohammad Elham Walizad, “Sign Language Recognition System using Convolutional Neural Network and Computer Vision”, “International Research Journal of Engineering and Technology (IRJET)”, 2020.