Sign language detection using CNN

Report in an application based project using python

partial fulfilment of the degree

Bachelor of Technology

in

Computer Science & Engineering

By

K.NIKHIL	**2303A51LB0**
N.SANJAY	**2303A51LA4**
K.NAGANISHITH	**2303A51L89**
G.VISHNU VARDHAN	**2303A51L87**
CH.ABHIRAM	**2303A51L84**

II YEAR CSE

UNDER THE GUIDENCE OF

B. SWATHI

Submitted to

DEPARTMENT OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE

SR University, Anantha Sagar, Hanamkonda

SR UNIVERSITY

INDEX

1.ABSTRACT

Sign language is the only tool of communication for the person who is not able to speak and hear anything. Sign language is a boon for the physically challenged peopleto express their thoughts and emotion. In this work, a novel scheme of sign languagerecognition has been proposed for identifying the alphabets and gestures in sign language. With the help of computer vision and neural networks we can detect the signs and give the respective text output.

Keyword's: Sign Language Recognition, Convolution Neural Network, Image Processing, Edge Detection, Hand Gesture Recognition.

2.INTRODUCTION

Speech impaired people use hand signs and gestures to communicate. Normal people face difficulty in understanding their language. Hence there is a need of a system which recognizes the different signs, gestures and conveys the information tothe normal people. It bridges the gap between physically challenged people and normal people.

2.1 SIGN LANGUAGE

It is a language that includes gestures made with the hands and other body parts, including facial expressions and postures of the body. It used primarily by peoplewho are deaf and dumb. There are many different sign languages as, British, Indian and American sign languages. British sign language (BSL) is not easily intelligible tousers of American sign Language (ASL) and vice versa.

A functioning signing recognition system could provide a chance for the inattentive communicate with non-signing people without the necessity for an interpreter. It might be wont to generate speech or text making the deaf more independent. Unfortunately, there has not been any system with these capabilities thusfar. during this project our aim is to develop a system which may classify signing accurately.

American Sign Language (ASL) is a complete, natural language that has the same linguistic properties as spoken languages, with grammar that differs from English. ASL is expressed by movements of the hands and face. It is the primary language of many North Americans who are deaf and hard of hearing and is used bymany hearing people as well.

3.MOTIVATION

The 2011 Indian census cities roughly 1.3 million people with “hearing impairment”. In contrast to that numbers from Indian National Association of the Deaf estimates that18 million people roughly 1 percent of Indian population are deaf. These statistics formed the motivation for our project. As these speech impairment and deaf people need a proper channel to communicate with normal people there is a need for a system. Not all normal people can understand sign language of impaired people. Our project hence is aimed at converting the sign language gestures into text that is readable for normal people.

4.PROBLEM STATEMENT

Speech impaired people use hand signs and gestures to communicate.

Normal people face difficulty in understanding their language. Hence there is a need of a system which recognizes the different signs, gestures and conveys the information to the normal people. It bridges the gap between physically challenged people and normal people.

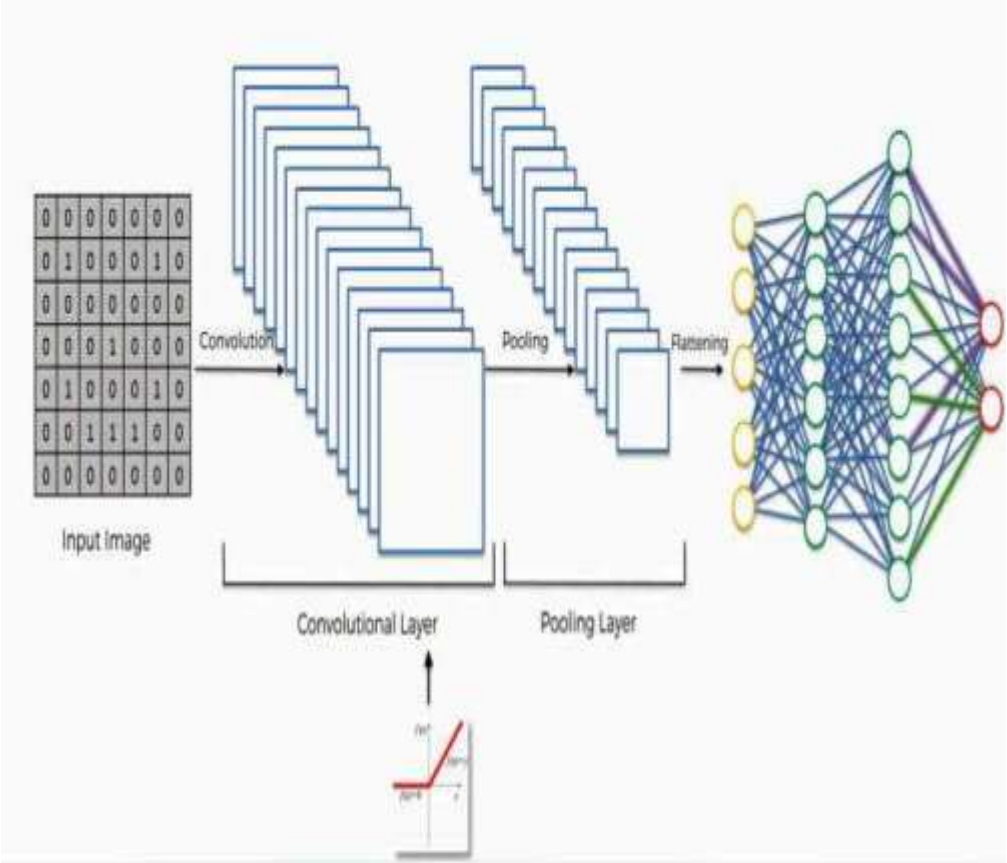
5.Convolution neural network algorithm:

Convolutional neural networks (CNN) is a special architecture of artificialneural networks, proposed by Yann LeCun in 1988. CNN uses some features of thevisual cortex. One of the most popular uses of this architecture is image classification.For example Facebook uses CNN for automatic tagging algorithms, Amazon for generating product recommendations and Google for search through among photos,

Instead of the image, the computer sees an array of pixels. For example, if image size is 300 x 300. In this case, the size of the array will be 300x300x3. Where 300 is width, next 300 is height and 3 is RGB channel values. The computer is assigned a value from0 to 255 to each of these numbers. value describes the intensity of the pixel at each point.

To solve this problem the computer looks for the characteristics of the baselevel. In human understanding such characteristics are for example the trunk or large ears. For the computer, these characteristics are boundaries or curvatures. And then through thegroups of convolutional layers the computer constructs more abstract concepts. In moredetail: the image is passed through a series of convolutional, nonlinear, pooling layersand fully connected layers, and then generates the output.

Comparing all the collected data with already existing data in the database to match aface with a name. A similar process is followed for scene labeling as well.Analyzing Documents: Convolutional neural networks can also be used for document analysis. This is not justuseful for handwriting analysis, but also has a major stake in recognizers. For a machine to be able to scan an individual's writing, and then compare that to the wide database it has, it must execute almost a million commands a minute. It is said with theuse of CNNs and newer models and algorithms, the error rate has been brought down toa minimum of 0.4% at a character level, though it's complete testing is yet to be widelyseen.



Figure[5]

6.METHODOLOGY

6.1 TRAINING MODULE:

Supervised machine learning: It is one of the ways of machine learning where the model is trained by input data and expected output data. create such model, it is necessary to go through the following phases:

1. model construction
2. model training
3. model testing
4. model evaluation

6.2 Model construction: It depends on machine learning algorithms. In this projects case, it was neural networks. Such an algorithm looks like:

1. begin with its object: model = Sequential()
2. then consist of layers with their types: model. Add(type_of_layer())
3. after adding a sufficient number of layers the model is compiled. At this moment Kara's communicates with TensorFlow for construction of the model. During model compilation it is important to write a loss function and an optimizer algorithm. It looks like: model. compile(loss= 'name_of_loss_function', optimizer= 'name_of_opimazer_alg') The loss function shows the accuracy of each prediction made by the model.

Before model training it is important to scale data for their further use.

6.3 Model training:

After model construction it is time for model training. In this phase, the model is trained using training data and expected output for this data. look s this way: model. Fit(training data, expected output). Progress is visible on the console when thescript runs. At the end it will report the final accuracy of the model.

6.4 Model Testing:

During this phase a second set of data is loaded. This data set has never been seen by the model and therefore true accuracy will be verified. After the model training is complete, and it is understood that the model shows the right result it can be saved by; Finally, the saved model can be used inthe real world. The name of this phase is model evaluation. This means that the modelcan be used to evaluate new data.

6.5 IMAGE PROCESSING

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too.

Image processing basically includes the following three steps:

- Importing the image via image acquisition tools.
- Analyzing and manipulating the image.
- Output in which result can be altered image or report that is based on image analysis.

There are two types of methods used for image processing namely, analogue and digital image processing. Analogue image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in manipulation of the digital images by using computers. The three general phases that all types of data have to undergo while using digital technique are pre- processing, enhancement, and display, information extraction.

Digital image processing:

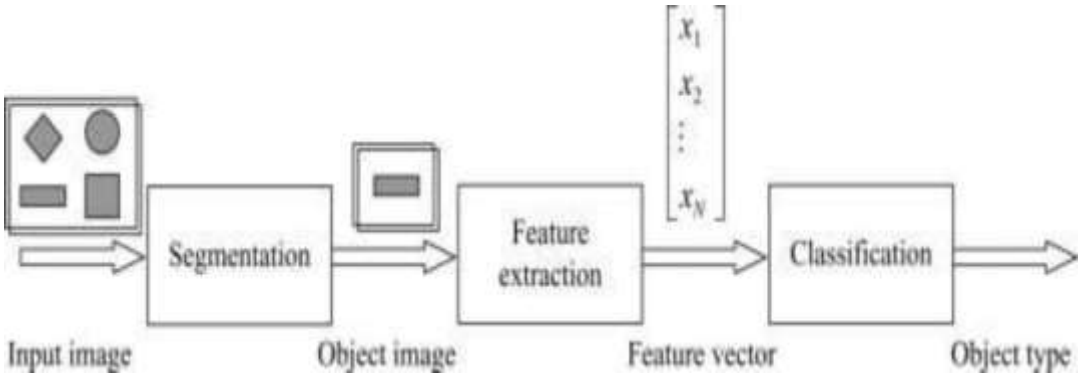
Digital image processing [1] consists of the manipulation of images using digital computers. Its use has been increasing exponentially in the last decades. Its applications range from medicine to entertainment, passing by geological processing and remote sensing. Multimedia systems, one of the pillars of the modern informationsociety, rely heavily on digital

image processing.

Digital image processing consists of the manipulation of those finite precisionnumbers. The processing of digital images can be divided into several classes: imageenhancement, image restoration, image analysis, and image compression. In image enhancement, an image is manipulated, mostly by heuristic techniques, so that a human viewer can extract useful information from it.

Digital image processing is to process images by computer. Digital image processing can be defined as subjecting a numerical representation of an object to a series of operations in order to obtain a desired result. Digital image processing consists of the conversion of a physical image into a corresponding digital image andthe extraction of significant information from the digital image by applying various algorithms.

Pattern recognition: On the basis of image processing, it is necessary to separate objects from images by pattern recognition technology, then to identify and classify these objects through technologies provided by statistical decision theory. Under the conditions that an image includes several objects, the pattern recognition consists of three phases, as shown in Fig.



Fig[7]: Phases of pattern recognition

The first phase includes the image segmentation and object separation. In thisphase, different objects are detected and separate from other background. The secondphase is the feature extraction. In this phase, objects are measured. The measuring feature is to quantitatively estimate some important features of objects, and a group ofthe features are combined to make up a feature vector during feature extraction. The third phase is classification. In this phase, the output is just a decision to determine which category every object belongs to. Therefore, for pattern recognition, what inputare images and what output are object types and structural analysis of images. The structural analysis is a description of images in order to correctly understand and judge for the important information of images

6.6 SEGMENTATION

Image segmentation is the process of partitioning a digital image into multiple segments(sets of pixels, also known as image objects). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse. Modern image segmentation techniques are powered by deep learning technology. Here are several deep learning architectures used for segmentation:

Why does Image Segmentation even matter?

If we take an example of Autonomous Vehicles, they need sensory input devices like cameras, radar, and lasers to allow the car to perceive the world around it, creating a digital map. Autonomous driving is not even possible without object detection which itself involves image classification/segmentation.

How Image Segmentation works

Image Segmentation involves converting an image into a collection of regions of pixels that are represented by a mask or a labeled image. By dividing an image into segments, you can process only the important segments of the image instead of processing the entire image. A common technique is to look for abrupt discontinuities in pixel values, which typically indicate edges that define a region.Another common approach is to detect similarities in the regions of an image. Some techniques that follow this approach are region growing, clustering, and thresholding. A variety of other approaches to perform image segmentation have been developed over the years using domain-specific knowledge to effectively solve segmentation problems in specific application areas.

6.7 CLASSIFICATION: CONVOLUTION NEURAL NETWORK

[4]Image classification is the process of taking an input(like a picture) and outputting its class or probability that the input is a particular class. Neural networks are applied in the following steps:

One hot encode the data: A one-hot encoding can be applied to the integer representation. This is where the integer encoded variable is removed and a new binary variable is added for each unique integer value.

1) Define the model: A model said in a very simplified form is nothing but a function that is used to take in certain input, perform certain operation to its beston the given input (learning and then predicting/classifying) and produce the suitable output.
2) Compile the model: The optimizer controls the learning rate. We will be using 'adam' as our optmizer. Adam is generally a good optimizer to use for many cases. The adam optimizer adjusts the learning rate throughout training. The learning rate determines how fast the optimal weights for the model are calculated. A smaller learning rate may lead to more accurate weights (up to a certain point), but the time it takes to compute the weights will be longer.
3) Train the model: Training a model simply means learning (determining) good values for all the weights and the bias from labeled examples. In supervised learning, a machine learning algorithm builds a model by examining many examples and attempting to find a model that minimizes loss; this process is called empirical risk minimization.
4) Test the model

A convolutional neural network convolves learned featured with input data and uses 2D convolution layers.

6.7.1 Relu Layer:

Rectified linear unit is used to scale the parameters to non negativevalues.We get pixel values as negative values too . Inthis layer we make them as 0's.The purposeof applying the rectifier function is to increase the non-linearity in our images. The reason we want to do that is that images are naturally non-linear. The rectifier serves to break up the linearity even further in order to make up for the linearity that we might impose an image when we put it through the convolution operation. What the rectifier function does to an image like this is remove all the black elements from it, keeping only those carrying a positive value (the grey and white colors).The essentialdifference between the non-rectified version of the image and the rectified one is the progression of colors.

After we rectify the image, you will find the colors changing more abruptly. The gradual change is no longer there. That indicates that the linearityhas been disposed of.

6.7.2 Pooling Layer:

The pooling (POOL) layer reduces the height and width of the input. It helps reduce computation, as well as helps make feature detectors more invariant to its position in the input This process is what provides the convolutional neural network with the "spatial variance" capability. In addition to that, pooling serves to minimize the size of the images as well as the number of parameters which, in turn, prevents an issue of "overfitting" from coming up. Overfitting in a nutshell is when you create an excessively complex model in order to account for the idiosyncracies we just mentioned The result ofusing a pooling layer and creating down sampled or pooled feature maps is a summarized version of the features detected in the input. They are useful as small changes in the location of the feature in the input detected by the convolutional layer will result in a pooled feature map with the feature in the same location. Thiscapability added by pooling is called the model's invariance to local translation.

6.7.3 Fully Connected Layer:

The role of the artificial neural network is to take this data and combine the features into a wider variety of attributes that make the convolutional network more capable of classifying images, which is the whole purpose from creating a convolutional neural network. It has neurons linked to each other ,and activates if it identifies patterns and sends signals to output layer .the outputlayer gives output class based on weight values, For now, all you need to know is that the loss function informs us of how accurate our network is, which we then use in optimizing our network in order to increase its effectiveness. That requires certain things to be alteredin our network. These include the weights (the blue lines connecting the neurons, which are basically the synapses), and the feature detector since the network often turns out to be looking for the wrong features and has to be reviewed multiple times for the sake of optimization.This full connection process practically works as follows:

- The neuron in the fully-connected layer detects a certain feature; say, a nose.
- It preserves its value.
- It communicates this value to the classes trained images.

7.EXISTING SYSTEM

In Literature survey we have gone through other similar works that are implemented in the domain of sign language recognition. The summaries of each of the project works are mentioned below

A Survey of Hand Gesture Recognition Methods in Sign LanguageRecognition

[1]Sign Language Recognition (SLR) system, which is required to recognize sign languages, has been widely studied for years. The studies are based on various input sensors, gesture segmentation, extraction of features and classification methods. This paper aims to analyze and compare the methods employed in the SLR systems, classifications methods that have been used, and suggests the most promising method for future research. Due to recent advancement in classification methods, many of the recent proposed works mainly contribute on the classification methods, such as hybridmethod and Deep Learning. This paper focuses on the classification methods used in prior Sign Language Recognition system. Based on our review, HMM- based approaches have been explored extensively in prior research, including its modifications.

This study is based on various input sensors, gesture segmentation, extraction of features and classification methods. This paper aims to analyze and compare the methods employed in the SLR systems, classifications methods that have been used, and suggests the most reliable method for future research. Due to recent advancementin classification methods, many of the recently proposed works mainly contribute to the classification methods, such as hybrid method and Deep Learning. Based on our review, HMM-based approaches have been explored extensively in prior research, including its modifications. Hybrid CNN-HMM and

fully Deep Learning approaches have shown promising results and offer opportunities for further exploration.

Communication between Deaf-Dumb People and Normal People

[2]Chat applications have become a powerful media that assist people to communicate indifferent languages with each other. There are lots of chat applications that areused different people in different languages but there are not such a chat application thathas facilitated to communicate with sign languages. The developed system is based on Sinhala Sign language. The system has included four main components as text messages are converted to sign messages, voice messages are converted to sign messages, sign messages are converted to text messages and sign messages are converted to voice messages. Google voice recognition API has used to develop speech character recognition for voice messages. The system has been trained for the speech and text patterns by using some text parameters and signs of Sinhala Sign language is displayed by emoji. Those emoji and signs that are included in this system will bring the normal people more close to the disabled people. This is a 2-way communication system but it uses pattern of gesture recognition which is not very reliable in getting appropriate output.

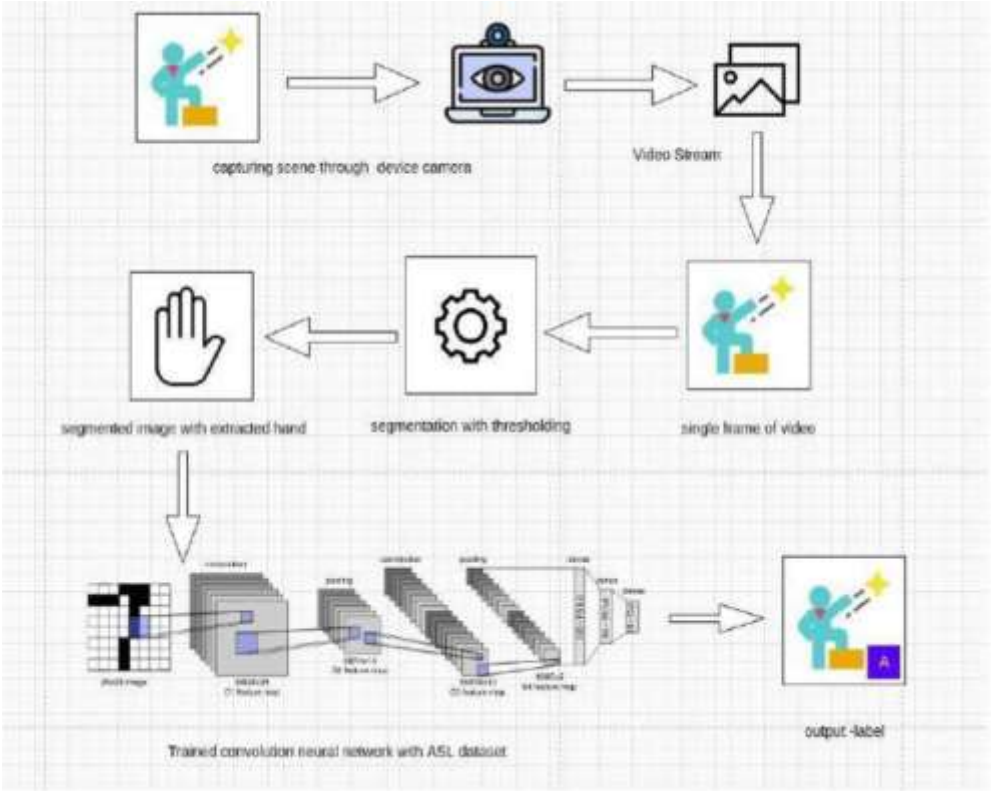
A System for Recognition of Indian Sign Language for Deaf People using Otsu's algorithm

[3]In this paper we proposed some methods, through which the recognition of the signs becomes easy for peoples while communication. And the result of those symbols signswill be converted into the text. In this project, we are capturing hand gestures throughwebcam and convert this image into gray scale image. The segmentation of gray scaleimage of a hand gesture is performed using Otsu thresholding algorithm. Total imagelevel is divided into two classes one is hand and other is background. The optimal threshold value is determined by computing the ratio between class variance and totalclass variance. To find the boundary of hand gesture in image Canny edge detection technique is used. In Canny edge detection we used edge-based segmentation and threshold-based segmentation. Then algorithm is used because of its simplecalculation and stability. This algorithm fails, when the global distribution of the targetand background vary widely.

8.PROPOSED SYSTEM

Our proposed system is sign language recognition system using convolution neural networks which recognizes various hand gestures by capturing video and converting it into frames. Then the hand pixels are segmented and the image it obtained and sent for comparison to the trained model. Thus our system is more robustin getting exact text labels of letters.

9.System Architecture



Figure[6]

10.SYSTEM CONFIGURATION

Software requirements

OperatingSystem: Windows, Mac, Linux

SDK: OpenCV, TensorFlow, Keros, Numpy

Hardware Requirements

The Hardware Interfaces Required are:

Camera: Good quality,3MP

Ram: Minimum 8GB or higher

GPU: 4GB dedicated

Processor: Intel Pentium 4 or Higher

HDD: 10GB or higher

Monitor:15” or 17” color monitor

Mouse: Scroll or Optical Mouse or Touch Pad

Keyboard: Standard 110 keys keyboard

11.SCREENSHOTS OF RESULTS

Fig a: Gesture 1 showing Warning text

Fig a: Gesture 1 showing Warning text

Fig c: Gesture 3 showing Please help me

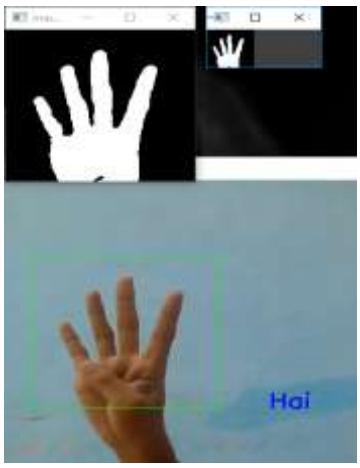


Fig d: Gesture 4 showing Hai text

Fig e: Gesture 5 showing stop text

Fig f: Gesture 6 showing Good text

Fig h: Gesture 8 showing Rock text

Fig g: Gesture 7 How are you? text

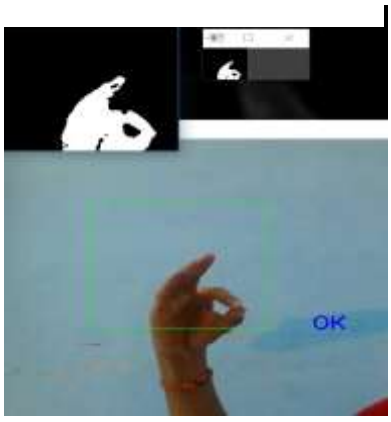


Fig i : Gesture 9 showing ok text

Fig j: No gesture showing do nothing

12.CONCLUSION AND FUTURE SCOPE

Nowadays, applications need several kinds of images as sources of informationfor elucidation and analysis. Several features are to be extracted so as to perform various applications. When an image is transformed from one form to another such as digitizing, scanning, and communicating, storing, etc. degradation occurs. Therefore, the output image has to undertake a process called image enhancement, whichcontains of a group of methods that seek to develop the visual presence of an image. Image enhancement is fundamentally enlightening the interpretability or awareness ofinformation in images for human listeners and providing better input for other automatic image processing systems. Image then undergoes feature extraction using various methods to make the image more readable by the computer. Sign language recognition system is a powerful tool to prepare an expert knowledge, edgedetect and the combination of inaccurate information from different sources.

The intend of convolution neural network is to get the appropriate classification

13.Future work

The proposed sign language recognition system used to recognize sign language text and can be further extended to recognize gestures facial expressions. Instead of displaying letter labels it will be more appropriate to display sentences as more appropriate translation of language. This also increases readability. The scope ofdifferent sign languages can be increased. More training data can be added to detect the letter with more accuracy. This project can further be extended to convert the signsto speech.

14.References:

[1] Suharjito Suharjito," A Survey of Hand Gesture Recognition Methods in Sign Language Recognition" Pertanika Journal of Science and Technology, volume26(4):1659-1675(pages), October 2018

[2] Chandrapal U. Chauhan, Minakshi Meshram, Samiksha Borkar, Vrushali Khobragade, Prasad Patil "Communication between Deaf-Dumb People and Normal People: SILENCE SPEAK "international journal of creative research thoughts, | Volume 10. 2320-2882(pages),2022

[3] Ms. Manisha D. Raut, , Ms. Pallavi Dhok , Mr. Ketan Machhale, Ms. Jaspreet Manjeet Hora," A System for Recognition of Indian Sign Language for Deaf People using Otsu's Algorithm", International Research Journal of Engineering and Technology (IRJET), Volume: 02, Apr-2015

[4] https://adeshpande3.github.io/

[5]https://miro.medium.com/v2/resize:fit:1358/1*8vLa14xeEN0Wbspp0fS-Bw.png

[6]https://www.ijraset.com/research-paper/sign-language-recognition-system-using-neural-networks

[7]https://d3i71xaburhd42.cloudfront.net/875bdb988964dffd887b52de47cc7803cbdd5f5/2-Figure1-1.png