International Journal of Advanced Engineering Science and Information Technology (IJAESIT)
ISSN: 2349-3216
Vol.6, No.6, June 2023

Sign Language Recognition using Machine Learning

Vishnupiriyan S
B.Tech – Information Technology
Velammal College of Engineering and Technology
vishnupiriyan2002@gmail.com

Darshan Ajith K R
B.Tech – Information Technology
Velammal College of Engineering and Technology
darshanajith02@gmail.com

Abdul Rahuman R
B.Tech – Information Technology
Velammal College of Engineering and Technology
rahumanabdul1806@gmail.com

Gobinath A
Assistant Professor/IT
Velammal College of Engineering and Technology
gopigopinath10@gmail.com

Abstract
Sign language is a computer vision-based complete convoluted language that engrosses signs shaped by the movements of hands in combination with facial expressions. It is a natural language used by people with low or no hearing sense for communication. Sign language, as a different form of the communication language, is important to large groups of people in society. There are different signs in each sign language with variability in hand shape, motion profile, and position of the hand, face, and body parts contributing to each sign. So, visual sign language recognition is a complex research area in computer vision. Sign language for communication is efficacious for humans, and vital research is in progress in computer vision systems. The earliest work in Indian Sign Language (ISL) recognition considers the recognition of significant differentiable hand signs and therefore often selecting a few signs from the ISL for recognition. This paper deals with robust modelling of static signs in the context of sign language recognition using deep learning-based convolutional neural networks (CNN). The project makes various voice controlled virtual assistants respond to hand gestures and also produces results in form of text outputs. The system is developed the deep learning algorithm such as CNN (VGG19). The experimental results shows that, the predicted results will be converted to speech and stored in'.mp3' file.

Keywords– *Supervised learning technique, Feature extraction techniques, Fingre spelled sign learning.*

I. INTRODUCTION

Machine Learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values. It is important because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. Classical machine learning is often categorized by how algorithms learns to become more accurate in its predictions. There are four basic approaches: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. Nowadays, Virtual Assistant devices have been part and parcel of our lives, but most of them are Voice Automated. Most commonly used Virtual Assistants are Alexa, Google Home, Apple Siri and Microsoft Cortana. These assistants listen to user's queries and respond accordingly making there life easier, thus they have been a very important part of Home Automation. Since these assistants are purely Voice Automated, Deaf-Mutes find it hard to make use of such technology. The agenda of the project is to develop an interface that will help the Deafmutes to use these Virtual

Assistants easily with easy. As of now, it might seem irrelevant to design such a system but in a longer run it might help deaf-mutes to equally enjoy their social and personal life. Designing such an interface will make them find their freedom while using such technologies and might boost their confidence in this Digital Age. 2 This paper focuses on a research that gives an idea of combining two modern technologies that are Hand Gesture Recognition and Virtual Voice Assistants in order to make it possible for people with hearing/speaking difficulties to interact with Digital Gadgets and also communicate with the outside world. Sign language is a computer vision- based complete convoluted language that engrosses signs shaped by the movements of hands in combination with facial expressions. It is a natural language used by people with low or no hearing sense for communication. A sign language can be used for communication of letters, words or sentences using different signs of the hands. This type of communication makes it easier for hearing-impaired people to express their views and also help in bridging the communication gap between hearing-impaired people and other person. Humans have been adapting to sign language to communicate since ancient times. Hand gestures are as ancient as the human civilization itself. Hand signs are especially useful to express any word or feeling to communicate. Therefore, people around the world use signals from hand constantly to express despite the formulation of writing conventions. In recent times, much research has been ongoing in developing systems that are able to classify signs of different sign languages into the given class. Such systems have found applications in games, virtual reality environments, robot controls and natural language communications. At present, the Indian Sign Language systems are in the developing stage and no sign language recognition system is available for recognizing signs in real time. So, there is a need to develop a complete recognizer which identifies signs of Indian Sign Language. The automatic recognition of human signs is a complex multidisciplinary problem that has not yet been completely solved. In the past years, a number of approaches were used which involve the use of machine learning techniques for sign language recognition. Since the advent of deep learning techniques, there have been attempts to recognize human signs. Networks which are based on deep learning paradigms deal with the architectures and learning algorithms that are biologically inspired, in distinction to conventional networks. Generally, the training of deep networks occurs in a layer-wise manner and depends on more distributed features as present in the human visual cortex. In this, the abstract features from the collected signs in the first layer are grouped into primary features in the second layer, which further combined into more defined features present in the next layer. These features are then further combined together into more engrossing features in the following layers, which help in the better recognition of different signs.

II. RELATED WORK

Input Image:
The dataset contains the images in the form of '.jpg' or '.png'. In this step, we have to read or load the input image by using the imread ()function. The input image is used to recognize the sign language. In our process, we are used the tkinter file dialogue box for selecting the input image.

Image preprocessing:
In our process, we have to resize the image and convert the image into grayscale. To resize an image, you call the resize () method on it, passing in a two- integer tuple argument representing the width and height of the resized image. The function doesn't modify the used image; it instead returns another Image with the new dimensions. Convert an Image to Grayscale in Python Using the Conversion Formula and the matplotlib Library. We can also convert an image to grayscale using the standard RGB to grayscale conversion formula that is imgGray = 0.2989 * R + 0.5870 * G + 0.1140 * B.

Feature extraction:
In our process, we can extract the features from pre-processed image by using mean standard deviation. Standard deviation is the spread of a group of numbers from the mean. Thevariance measures the average degree to which each point differs from the mean. While standard deviation is the square root of the variance, variance is theaverage of all data points within a group.

Image Splitting:
During the machine learning process, data are needed so that learning cantake place. In addition to the data required for training, test data are needed to evaluatethe performance of the algorithm in order to see how well it works. In our process, we considered 70% of the input dataset to be the trainingdata and the remaining 30% to be the testing data. Data splitting is the act of partitioning available data into two portions,usually for cross-validator purposes. 17

Classification:
In our process, we have to implement transfer learning such as VGG19. VGG-19 is a convolutional neural network that is 19 layers deep.You can load a pretrained version of the network trained on more than a million images from the Image Net database. The pretrained network can classify images into 1000 object categories,such as keyboard, mouse, pencil, and many animals.

Natural Language Processing:
NLP is a field in machine learning with the ability of a computer to understand, analyze, manipulate, and potentially generate human language. Cleaning (or pre-processing) the data typically consists of a number ofsteps:
• *Remove punctuation*: Punctuation can provide grammatical context to asentence which supports our understanding.

• *Tokenization*: Tokenizing separates text into units such as sentences orwords. It gives structure to previously unstructured text. eg: Plata o Plomo-

> 'Plata','o','Plomo'.
• *Stemming*: Stemming helps reduce a word to its stem form**.**

• *Sentiment analysis***:** In this step, we can analyse the sentiment into positive,neutral and negative by using the sentiment analyser (polarity score).

• Sentiment analysis works by breaking a message down into topic chunksand then assigning a sentiment score to each topic.

Result Generation:
The Final Result will get generated based on the overall classification and prediction. **Accuracy**:Accuracy of classifier refers to the ability of classifier. It predicts theclass label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.
AC= (TP+TN)/ (TP+TN+FP+FN)

III. THE PROPOSED MECHANISM

In proposed system, the Indian sign language dataset was taken as input. Then, we have to implement the pre-processing step. In this step, we have to resize the original image as well as grey scale conversion. After that, we have to extract the features from the pre-processed image by mean standard deviation. Then, we have to implement the image splitting such as test and train. Test is used for predicting and train is used for evaluating the model. After that, we have to implement the deep learning algorithm such as Convolutional Neural Network (VGG 19). The experimental results shows that some performance metrics such as accuracy, loss value and detect or classify the different types of sign language.

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.

Preprocessing:
An aspect ratio is a proportional relationship between an image's width and height. Essentially, it describes an image's shape. Aspect ratios are written as a formula of width to height, like this: For example, a square image has an aspect ratio of 1:1, since the height and width are the same. The image could be 500px × 500px, or 1500px × 1500px, and the aspect ratio would still be 1:1.As another example, a portrait-style image might have a ratio of 2:3. With this aspect ratio, the height is 1.5 times longer than the width. So the image could be 500px × 750px, 1500px × 2250px, etc. Cropping to an aspect ratio 12 aside from using built in

site style options , you may want to manually crop an image to a certain aspect ratio. For example, if you use product images that have same aspect ratio, they'll all crop the same way on your site.

Image scaling:

In compuer graphics and digital imaging , image scaling refers to the resizing of a digital image. In video technology, the magnification of digital material is known as upscaling or resolution enhancement. When scaling a vector graphic image, the graphic primitives that make up the image can be scaled using geometric transformations, with no loss of image quality. When scaling a raster graphics image, a new image with a higher or lower number of pixels must be generated. In the case of decreasing the pixel number (scaling down) this usually results in a visible quality loss. From the standpoint of digital signal processing, the scaling of raster graphics is a two-dimensional example of sample-rate conversion, the conversion of a discrete signal from a sampling rate (in this case the local sampling rate) to another.

Algorithm:

VGG19:

VGG19 is a variant of VGG model which in short consists of 19 layers (16 convolution layers, 3 Fully connected layer, 5 MaxPool layers and 1 SoftMax layer). There are other variants of VGG like VGG11, VGG16 and others. VGG19 has 19.6 billion FLOPs. A fixed size of (224 * 224) RGB image was given as input to this network which means that the matrix was of shape (224,224,3). The only preprocessing that was done is that they subtracted the mean RGB value from each pixel, computed over the whole training set. Used kernels of (3 * 3) size with a stride size of 1 pixel, this enabled them to cover the whole notion of the image. Spatial padding was used to preserve the spatial resolution of the image. Max pooling was performed over a 2 * 2 pixel windows with sride 2. This was followed by Rectified linear unit(ReLu) to introduce non-linearity to make the model classify better and to improve computational time as the previous models used tanh or sigmoid functions this proved much better than those. Implemented three fully connected layers from which first two were of size 4096 and after that a layer with 1000 channels 13 for 1000-way ILSVRC classification and the final layer is a softmax function.

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? 31 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 14 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.

Classification-CNN architecture:

The objective of CNN is to learn the features present in the data with higher order using convolutions. The CNN architecture works well for the recognition of objects which includes images. They can recognize individuals, faces, street signs and other facets of visual data. There exist a number of CNN variations, but each of them is based on the pattern of layers present. CNN architecture consists of different components which include different types of layers and activation functions. The listing describes the purpose and functioning of some commonly used layers which is discussed below. Convolutional layer. The core building blocks of CNN architecture are the convolutional layer. Convolutional layers (Conv) modify the input data with the help

of a patch of neurons connected locally from the previous layer. The dot product will be computed by the layer between the region of the neurons present in the input layer and the weights to which they are locally connected present in the output layer.

IV. PERFORMANCE EVALUATION

Input module:

Figure 1: Input Image

Preprocessing Module:

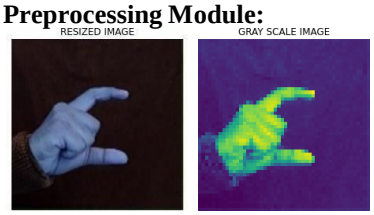

Figure 2: Preprocessing Image

Feature Extraction Module:

```
====================================
         Feature Extraction
====================================

[33.1724, 18.0, 1560.6338782399996]
```

Figure 3: Feature Extraction

Performance Module:

```
============================================================
---------- Convolutional Neural Network (VGG 19) ----------
============================================================


1.Accuracy is : 99.97278911620378 %

2.Loss is : 0.027210883796215057
```

Figure 4: Accuracy of the CNN

Prediction Module:

```
================================================
------------------ Prediction ------------------
================================================
--------------------------

Sign language = C

------------------------
```

Figure 5: Output is Predicted

```
------------------------
Text to speech conversion starting ...........


Text to speech conversion completed ...........
```

Figure 6: Output text is converted into Speech

V. CONCLUSION

We conclude that, the dataset was taken from dataset repository. We are extracted the features from pre- processed image by using mean standard variance. We are developed the Deep Learning algorithm such as VGG19. Finally, the experimental results shows that accuracy. Then, predicted results are converted to audio and then, audio is stored in .mp3 file.

REFERENCES

[1] Aman pathak,Avinash Kumar,Priyam,Priyam Guptha Chugh, "Real Time Sign Language Detection" International Journal for modern Trends in Science and Technology.

[2] Dr. Pallavi Chaudhari , Pranay Pathrabe , Umang Ghatbandhe , Sangita Mondal , Sejal Parmar, "Sign Language Detection System" International Journal of Engineering Technology and Management Science.

[3] Subhashini Yadav, Shreyashi Raj, Kashish Awasthi, Rohit Bidwan,"Real Time Sign Language Detection" International Journal for modern Trends in Science and Technology.

[4] Amey Chavan, Shubham Deshmukh, Favin Fernandes," Sign Language Detection" International Journal for modern Trends in Science and Technology.

[5] S Naga Parameswara Reddy, Chitrala Himavanth Sai Ram, Shaik Mansur, K.Pandiara," Sign Language Detection Using Deep Learning" International Journal for modern Trends in Science and Technology.

Authors Profile

My name is Vishnupiriyan. I am studying B.Tech Information Technology in Velammal College of Engineering and Technology.

My name is Darshan Ajith. I am studying B.Tech Information Technology in Velammal College of Engineering and Technology.

My name is Abdul Rahuman. I am studying B.Tech Information Technology in Velammal College of Engineering and Technology.

Myself is Gopinath. I have completed ME (Ph.D) and working as Assistant Professor/IT in Velammal College of Engineering and Technology.