SOLUTION ARCHITECTURE

Project: Real -Time communication System Powered By AI For Specially Abled

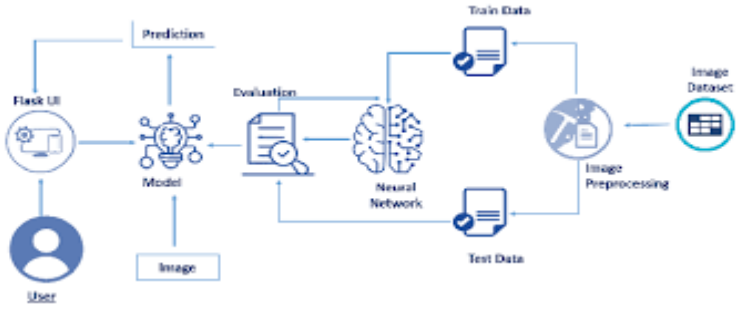
Team Members: Sathish P, Karthikeyan S, Nobi Prakash S, Jai Prakash K

PROJECT DESCRIPTION:

In our society, we have people with disabilities. The technology is developing day by day but no significant developments are undertaken for the betterment of these people. Communications between deaf-mute and a normal person has always been a challenging task. It is very difficult for mute people to convey their message to normal people. Since normal people are not trained on hand sign language. In emergency times conveying their message is very difficult. The human hand has remained a popular choice to convey information in situations where other forms like speech cannot be used. Voice Conversion System with Hand Gesture Recognition and translation will be very useful to have a proper conversation between a normal person and an impaired person in any language.

The project aims to develop a system that converts the sign language into a human hearing voice in the desired language to convey a message to normal people, as well as convert speech into understandable sign language for the deaf and dumb. We are making use of a convolution neural network to create a model that is trained on different hand gestures. An app is built which uses this model. This app enables deaf and dumb people to convey their information using signs which get converted to human-understandable language and speech is given as output.

TECHNICAL ARCHITECTURE:



SOLUTION:

HaGRID Dataset Description:

HaGRID (Hand Gesture Recognition Image Dataset) for hand gesture recognition (HGR) systems. This dataset contains 552,992 samples divided into 18 classes of gestures. The annotations consist of bounding boxes of hands with gesture labels and markups of leading hands. The proposed dataset allows for building HGR systems, which can be used in video conferencing services, home automation systems, the automotive sector, services for people with speech and hearing impairments, etc. We are especially focused on interaction with devices to manage them. That is why all 18 chosen gestures are functional, familiar to the majority of people, and may be an incentive to take some action. In addition, we used crowdsourcing platforms to collect the dataset and took into account various parameters to ensure data diversity. We describe the challenges of using existing HGR datasets for our task and provide a detailed overview of them. Furthermore, the baselines for the hand detection and gesture classification tasks are proposed.

Approach:

We will approach this project by using a three-layered Neural Network.

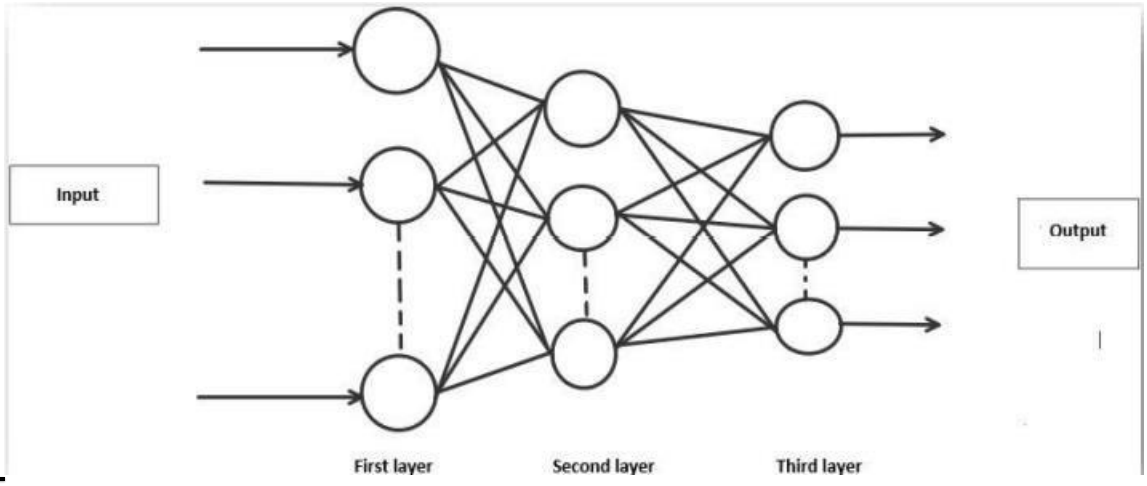
- The input layer: It distributes the features of our examples to the next layer for calculation of activations of the next layer.
- The hidden layer: They are made of hidden units called activations providing nonlinear ties for the network. A number of hidden layers can vary according to our requirements.
- The output layer: The nodes here are called output units. It provides us with the final prediction of the Neural Network on the basis of which final predictions can be made.

A neural network is a model based on how the brain functions. It is made up of several layers with numerous activations, which mirror neurons in our brain. A neural network attempts to learn a set of parameters from a set of

data, which may aid in recognising underlying links. Because neural networks can adapt to changing input, they can produce the best possible results without having to rethink the output criteria.

METHODOLOGY:

We created a Neural Network with one hidden layer and 100 activation units (excluding bias units). Data is loaded from a.mat file, then features (X) and labels (Y) are extracted. Then, to avoid overflow during computation, features are divided by 255 and rescaled into a range of [0,1]. The data is divided into 60,000 training instances and 10,000 testing examples.
Feedforward is used with the training set to calculate the hypothesis, followed by backpropagation to reduce the error between the layers. To solve the issue of overfitting, the regularisation parameter lambda is adjusted to 0.1. The optimizer is run 70 times to get the best fit model.



ALGORITHMS FOR HAND GESTURE RECOGNITION

MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. MATLAB provides the Toolboxes that allow you to learn and apply specialized technology. The recognition of hand gesture using MATLAB is described as follows. We use two algorithms for hand gesture recognition using MATLAB as Edge detection and Skin detection algorithms.

EDGE DETECTION

Following steps are used for detecting the edges:

1. Image capturing using a webcam or the front camera of the mobile phone.

2. Converting the captured image into frames.

3. Image pre-processing using Histogram Equalization.

4. Edge detection of the hand by using an algorithm like Canny Edge Detection.

5. Enlargement of the edges of regions of foreground pixels by using Dilation to get a continuous edge. 6. Filling of the object enclosed by the edge.

7. Storing the boundary of the object in a linear array.

8. Vectorization operation performed for every pixel.

9. Detection of the fingertips.

10. Tracking of the fingertips in consecutive frames to determine the motion.

WORKING:

- Neural Networks receive an input and transform it through a series of hidden layers.
- Each hidden layer is made up of a set of neurons, where each neuron is fully connected to

all neurons in the previous layer.

- Neurons in a single layer function completely independently.
- The last fully connected layer is called the "output layer“.

Convolution Layer:

The Convolutional layer is the core building block of a CNN. The layer's parameters consist of a set of learnable filters (or kernels), which have a small receptive field, but extend through the full depth of the input volume. During the forward pass, each filter is convolved across the width and height of the input volume, computing the dot product between the entries of the filter and the input and producing a 2- dimensional activation map of that filter. As a result, the network learns filters that activate when they see some specific type of feature at some spatial position in the input.

Feature Extraction:

All neurons in a feature share the same weights .In this way all neurons detect the same feature at different positions in the input image. Reduce the number of free parameters.

Subsampling Layer:

Subsampling, or down sampling, refers to reducing the overall size of a signal. The subsampling layers reduce the spatial resolution of each feature map. Reduce the effect of noises and shift or distortion invariance is achieved.

Pooling layer:

It is common to periodically insert a Pooling layer in-between successive Conv layer in a Convent architecture. Its function is to progressively reduce the spatial size of the representation to reduce the number of parameters and computation in the network, and hence to also control overfitting. The Pooling Layer operates independently on every depth slice of the input and resizes it spatially, using the MAX operation.

TensorFlow:

TensorFlow is an open-source machine learning library for research and production. TensorFlow offers APIs for beginners and experts to develop for desktop, mobile, web, and cloud. See the sections below to get started. By scanning the numerical digit and convert into png format using python3 command in terminal we can get text output and sound output.

RESULT:

As with any study or project undertaken in the field of machine learning and image recognition,

We do not consider our results to be perfect after processing. Machine learning is a field that is always evolving, and there is always room for advancement improvement in your process; there will always be something new strategy that produces superior outcomes for the same challenge The application has been submitted.

Three models were used: Multi-Layer Perceptron (MLP), Convolution Neural Network (CNN), and Network (CNN). The accuracy of the classifier varies depending on the model which demonstrates which is superior.