International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

MuteCom Using Tensorflow-Keras Model

Aniket Singh[1], Deepak Maurya[2], Shaunak Pacharkar[3], Prof. Bhanu Tekwani[4]

[1,2,3]Students of Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, Maharastra
[4]Prof. of Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, Maharastra

---

Abstract - *Communication is the main channel between people to communicate with each other. Since deaf and dumb people cannot communicate with normal person so they have to depend on of visual ways to communication. The problems faced by the deaf and dumb people at the present time and the difficulties of their communication with normal persons sparked our interest and led us to try to find a solution to their difficulties and to minimize them as much as possible. Because they represent a significant part of society and they need to deliver their ideas in the simplest way. So, our project aims to bridge this gap by enabling communication between dumb/deaf people on the one hand and normal people on other hand by introducing an application.*

This application uses in-build camera of the device to take input as the hand gesture of the person using the device. The application will recognize the hand sign and try to predict the actual meaning of that sign. Application uses multiple data sets which contains the hand gesture images and machine learning to predict the sign. The output of the prediction is given in the form of text/speech output so that the intended person does not feel any inconvenience in communication with that person.

Key Words: **CNN, American sign language database, tensorflow, google colab, Keras, OpenCV.**

1. INTRODUCTION

Nowadays we always hear about new technology that improves our lifestyle, that creates our life easier. Technology has revolutionized the human mankind. There are many languages spoken all round the world and interpreted. "Special people", people who have difficulty in speaking and hearing. Sometimes people interpret these messages wrongly either through signing or through lip reading or lip synchronization. This project is created in such the simplest way to assist these specially challenged people.

Gesture recognition plays important role in communication between the dumb/deaf people with normal person who are not able to understand the sign language made for special(deaf/dumb) people. The detection of hand gestures is often done using web camera. The pictures are then converted into standard size with the assistance of pre-processing. The aim of this project is to develop a system that can convert the hand gestures into text and the gives speech output. The focus of this project is to put the images within the database and with database matching the image is converted into text. The detection involves observation of hand movement and the trained model tries to predict what the gesture means and. The method gives output in text as well as speech format that helps to reduce the communication gap between deaf-mute and people.

1.1. Need: -

There has been less research for Deaf and Dumb people. This topic has got less attention as compared to other sectors. The main challenges that this special person(deaf/dumb) are facing is the communication gap between special person and normal person. Deaf and Dumb people always find it difficult to communicate with normal person. This huge challenge makes them uncomfortable and they feel discriminated in society. Because of miss communication deaf and dumb people feel not to communicate and hence they never able to express their feelings or the things the want to say. Hand Gesture Recognition to speech output system localizes and track the hand gestures of the dumb and deaf people in order to maintain a communication channel with the other people.

1.2. Objectives for making this project is as follows:

1. Easy communication for deaf people:

With the help of this application the hand gestures will be taken as an input and the output provided will be in both text and speech, the text output provided will help the deaf people to understand what the person is trying to say and make the communication between them easier.

2. Easy communication for dumb people:

With the help of this application the hand gestures will be taken as an input and the output provided will be in both text and speech, the speech output given will help the normal person understand the dumb individual to trying to say and this will make the communication between them easier.

3. Normal person can easily communicate with physically challenged people:

One can easily use our project with a basic knowledge of sign language and can easily communicate with specially challenged (deaf/dumb) people.

2. LITERATURE SURVEY

In Literature Review, we studied about existing project related to this topic and try to understand about existing system behaviour.

Shweta S. Shinde, Rajesh M. Autee and Vitthal K. Bhosale [1] have proposed a method in which the angle and peak calculation approach is used to extract the features of hand gestures by using MATLAB and then they convert the recognized gesture into speech using MATLAB inbuilt command.

Sangeetha. R.K, Valliammai .V and Padmavathi .S [2] have proposed a system based on the Indian hand sign language which contains both hands to create a gesture unlike the American sign language in which one hand is used. Their system is implemented using MATLAB without using any other external hardware for the user, here the runtime live image is captured after which image frames are extracted and image processing is applied using HIS model and then the feature extraction is done by distance transform method. The results obtained by this model is found to be satisfactorily good for most of the hand signs.

Anchal Sood and Anju Mishra [3] have proposed a sign recognition system based on Harris algorithm for extraction of feature in which after the image pre-processing part, the feature is extracted and stored in the Nx2 matrix. This matrix is further used to match the image from the database. There are some limitations to the system. The very light brown to somewhat dark brown background gives error as they are considered in the range value for skin segmentation. But the results are efficient.

Prashant G. Ahire, Kshitija B. Tilekar, Tejaswini A. Jawake, Pramod B. Warale [4] system works on MATLAB for hand gesture recognition, here they have taken a real time video as an input and after applying the image processing stage they have used the correlation based approach for mapping purpose and then at last the audio is generated using google TTS API. The system provides an efficient result as per the system is proposed.

Sonal Kumari and Suman K. Mitra [5] have proposed a system based on hand action recognition using background subtraction technique for image processing and they use Direct Fourier transform (DFT) algorithm for image` extraction based on the MATLAB.

3. Proposed System

This application takes input as hand gesture of the person and track the hand movement and simultaneously it converts the captured video input into the fames and it checks in database for those frames, if the image is already available then it gives the meaning of that hand gesture in the form of text as well as speech output. If the system is not able to find the gesture frame in database then it predicts that hand gesture from available hand gestures data set or images which are already present in the database.

Some features of the system are as follows: -

Hand Detection: - This application will only take the hand movement not the other part of the hand or body. So, the user should keep their hand in proper position so that the application can recognize the movement of the hand and give output according to that.

Prediction: - It might happen that some hand gesture or sign may not be available in the system's database in that case this application tries to predict that hand gesture or sign from available hand gesture or sign in the database.

Speech and Text Output: - This application gives feature of both the text as well as speech output. User can choose one or both as per their convenience.

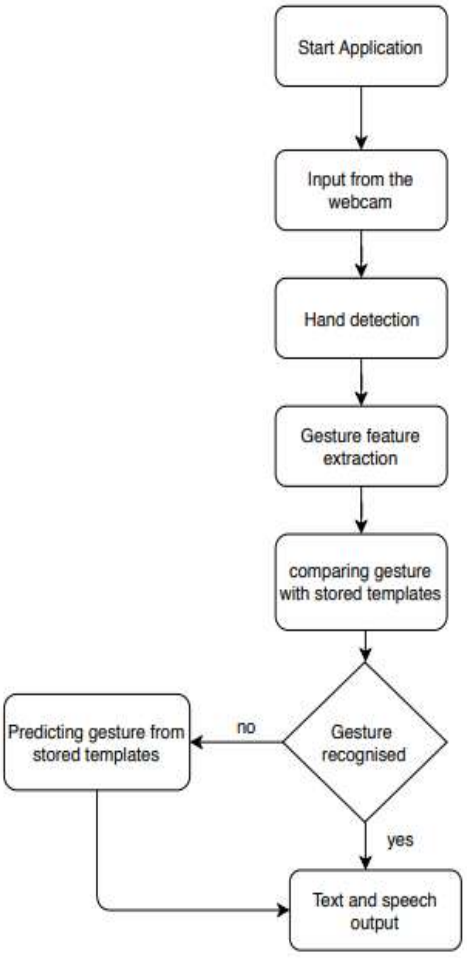

Fig-1: Working of MuteCom

4. Hardware and Software requirements:

Hardware Requirements: -
Minimum requirement - Laptop with configuration of 4GB RAM, 500GB Hard Disk, any OS and with a clearly working webcam is used in development of the application.
Software Requirements:

- Python, OpenCV, Tensorflow and Keras
- ASL database: Kaggle
- Training model: Google colab
- Version Control: Git and GitHub

5. Methodology

Neural Networks: - Neural networks are composed of straightforward elements operating in parallel. Neural networks are models that are capable of machine learning and pattern recognition. They are usually presented as systems of interconnected "neurons" which will compute values from inputs by feeding information through the network. Commonly neural networks are adjusted, or trained, in order that a selected input results in a specific target output. There, the network is adjusted, supported a comparison of the output and therefore the target, until the network output matches the target. Typically, many such input/target pairs are used, during this supervised learning, to train a network. Neural networks are trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech and vision and control systems.

TensorFlow: - The primary software tool of deep learning is TensorFlow. It is an open source AI library, using data flow graphs to create models. It allows developers to make largescale neural networks with many layers. TensorFlow is especially used for: Classification, Perception, Understanding, Discovering, Prediction and Creation.

Keras: - Keras is an open-source neural-network library written in Python. it's capable of running on top of TensorFlow. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible.

Image Database: - The starting point of the project was the creation of a database with all the images that would be used for training and testing. The image database can have different formats. Images can be either hand drawn, digitized photographs or a 3D dimensional hand. Photographs were used, as they are the most realistic approach. We needed a dataset with images of hands showing all the signs of ASL (American Sign Language) language. Creating a dataset with these images would require a lot of time therefore we obtained the dataset from Kaggle [7]. The dataset we obtained contained folders 29 with altogether 87000+ images as that would decide the robustness of the algorithm. We required a huge dataset to obtain more accurate and therefore, it had to be done in such a way that different situations could be tested.

Google Colab:- The model we created had be be trained on huge database which contained 87000+ images and hence to do so we required a huge amount of RAM and even on a laptop of 8GB RAM it was hard to train the model so to overcome this issue we used google colab[6] a cloud platform provided by google which provided us a GPU with 25GB RAM and through it we were able to train the model successfully with minimum time. Google Colab also helped us in overcoming the problem of training the model multiples when we wanted to use it. We trained the model on colab and obtained the .h5 and .json file which contained the final weights and hierarchy respectively and this helped in avoiding to train the model every time when we use it.

5.1. Implementation includes the following steps: -

1. Obtained a dataset from kaggle which had overall 87000+ images and 29 folders which included the signs of all the 26 alphabets of ASL language and the rest 3 folders contained the gestures of space, delete and nothing.

2. We created a model using tensorflow, keras and cv2 which we had to train on the database to obtain desired results.
3. We trained on google colab a cloud platform provided by google. Colab provided us with a GPU of 25GB of RAM which helped us in training our model with minimum time.

4. To avoid training the model multiple times when we want to use it, we trained the model once on google colab and downloaded the model.h5 and model.json files from there, were model.h5 file contains the final weights of the trained model which helps in obtaining the desired output and model.json file contained the hierarchy of execution. 5. We integrated these model.h5 and model.json files to our model stored in memory of our laptops using keras which helped is obtaining the output quickly by training it only once on colab.

6. The model works in the following manner: -

a. As we run the application it opens the webcam for taking the input of hand gestures.

b. Each and every frame is captured by the webcam and then extraction of its features takes place, which includes conversion of image into matrix, binarization etc.

c. After extraction of features, it fed to our pre-trained model for classifying the input gesture. When the gesture is matched successfully the output is printed on the screen. The model built provides an accuracy of approximately 99.67.

3. CONCLUSIONS

The accuracy of the model is 67%, accuracy can be increased if we train the model with real world data so that it can handle the real-world environment with more accuracy.

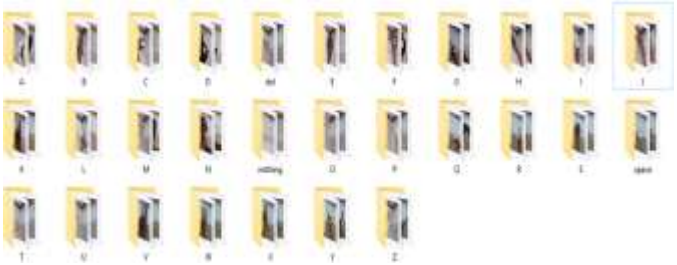

Fig-2: Folder structure of dataset

Fig-3: Folder structure for output of pre-trained model

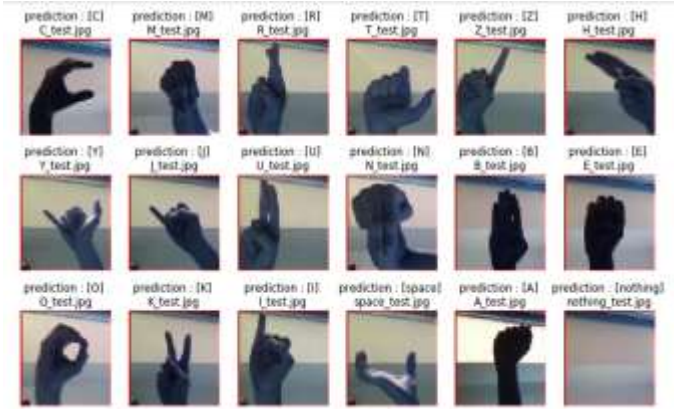

Fig-4:. Output of model for test data set:-

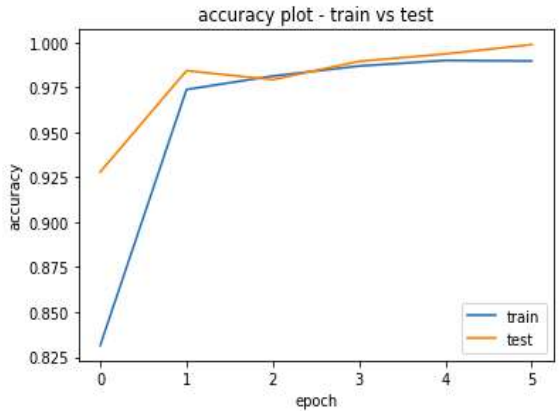

Fig-5: Accuracy plot of training vs test data

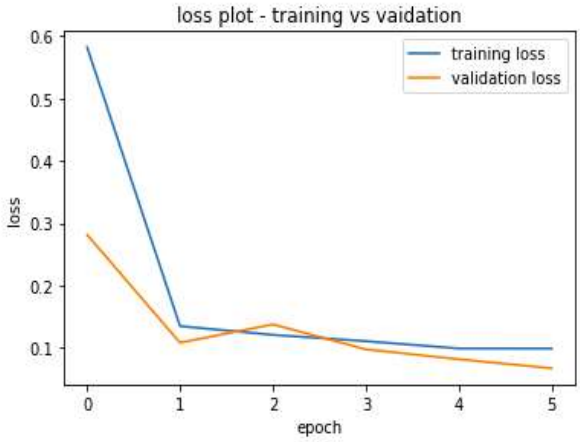

Fig-6: Loss plot of training vs loss

After testing the model, we have found that it is working correctly as desired in most of the cases.

REFERENCES

[1] Shinde, Shweta S., Rajesh M. Autee, and Vitthal K. Bhosale. "Real-time two-way communication approach for hearing impaired and dumb person based on image processing." Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on. IEEE, 2016.

[2] Shangeetha, R. K., V. Valliammai, and S. Padmavathi. "Computer vision-based approach for Indian Sign Language character recognition." Machine Vision and Image Processing (MVIP), 2012 International Conference on. IEEE, 2012.

[3] Sood, Anchal, and Anju Mishra. "AAWAAZ: A communication system for deaf and dumb." Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2016 5th International Conference on. IEEE, 2016.

[4] Ahire, Prashant G., et al. "Two Way Communicator between Deaf and Dumb People and Normal People." Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on. IEEE, 2015

[5] Kumari, Sonal, and Suman K. Mitra. "Human action recognition using DFT." Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011 Third National Conference on. IEEE, 2011.

[6] https://colab.research.google.com/

[7] https://www.kaggle.com/grassknoted/asl-alphabet