IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11118

Real Time Hand Gesture Recognition using Open CV and Convolutional Neural Network

Ansari Arbaz[1], Ankur Singh[2], Khan Anas[3], Prof. V. P. Tonde[4]

Student, IT, Sinhgad Institute of Technology, Lonavala, India[1-3]

Professor, IT, Sinhgad Institute of Technology, Lonavala, India[4]

Abstract: Computer Vision and deep learning techniques to recognize the hand gestures are among the trending domain of research now a days. The power of Artificial intelligence to improve the user interface and HCI is making human life much easier. Many researches are going on to develop systems that can understand hand gestures as input and perform several tasks. The communication through sign language is very ambiguous as it differs from person to person. This makes it very specific. Therefor, this project aims to build a system that can effectively determine a set of gestures, convert it to text and audio then perform certain task. At the same time it allows user to teach the system, their own gestures and associated messages to recognize.To accomplish this a CNN model is built to classify the gestures and Open CV is used for image capture and processing. After the model identifies the gesture it is converted to text/audio and associated task is performed.

Keywords: Computer Vision, Convolution Neural Network, Deep Learning,Tensor Flow, Keras, Tkinter

I. INTRODUCTION

Hand gestures are a form of non-verbal communication used in several fields such as communicating with deaf and mute people, robot control (Gesture Control Robotics), HCI, home automation, and medical applications. The project explains the possible solutions and efficient ways of communication using hand gestures. There are many forms of sign language that use different gestures to express one's needs and thoughts. These make communication very ambiguous. The system uses a VGG16 neural network model trained to predict hand gestures. But it also allows users to add their desired hand gestures and associated labels to the database, and then layers of the neural network are retrained using transfer learning. This neural network model is then used to convert the user's gestures into text or audio or perform any task. Open CV algorithms are used for data prepocessing.and CNN is used for building the classification model.

.

II. PROPOSED METHOD

A. System Description: The system used in this project can be described as multi layer perceptron. The below block diagram explains the different blocks of our project..

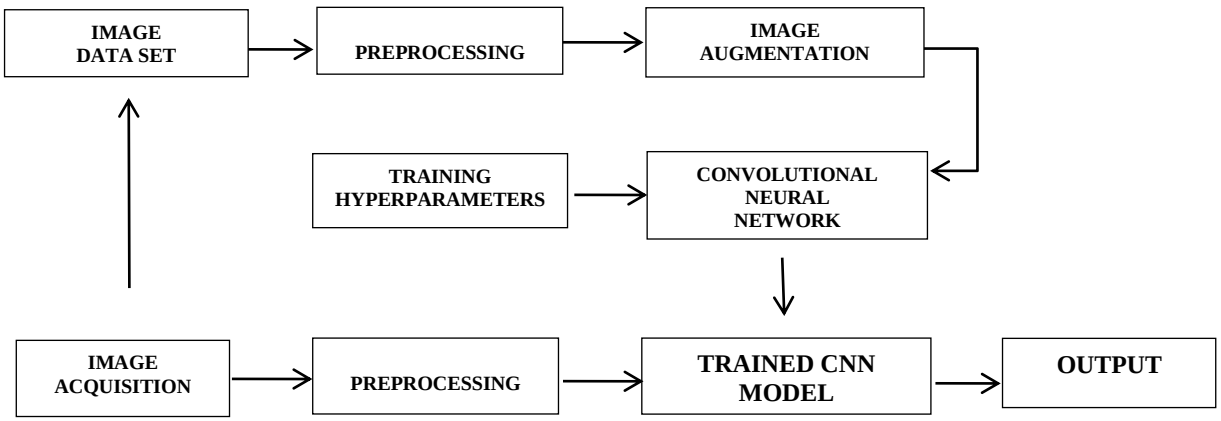

Fig. 1 System Block Diagram

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11118

A. Data Base:

The data base contains two image data sets. One of the data set contains images of hand gestures that convey simple text messages which are supposed to be converted into text and audio.

While the other data base contains images of gestures that command to draw some geometrical figures like triangle,rectangle and circle etc. Using these images commands of drawing associated shape are supposed to be done.

B. Image Preprocessing:

Image processing is done to remove the noise from data. Training the model over original images does not give sufficient accuracy. Image processing techniques such as converting from RGB to gray scale can reduce the training time. Different techniques can be applied to remove background and unwanted pixel values from the images.

Image Augmentation:

Data augmentation is technique to generate multiple copies of data from original data by applying some modifications. It is done to increase the amount and variety of data. Multiple operations such as Rotation, Mirroring, Cropping , Shearing Color shifting etc can be applied over images.

C. CNN Training:

Deep learning is used to train the model.Training options such as number of epoch, batch size, learning rate are set before fitting the model over data set. The neural network is then trained over the data base.

D. Image Acquisition:

Any camera or webcam can be used to capture the image. After processing the image it will reduced according to input size of neural network, hence high quality of camera is not needed to capture the image.If the user wants to train the system for new gesture image and label are modified in data base else is sent for classification.

E. Output:

The output of system is displayed in the form of text and audio. If the gesture commands for some task, it is performed e.g. Drawing a geometric shape.

III. IMPLEMENTATION

A. Data Base:

The data base contains two image data sets.

- Data base for gesture transaltion:

This data set contains images of hand gestures that convey simple text messages which are supposed to be converted into text and audio.The admin has trained the neural network over some images to classify them into respective labels. User can add their own hand gestures with associated message to this data base and the model learns the parameters of new data automatically.

- Data base for drawing geometric shapes:

While the other data base contains images of gestures that command to draw some geometrical figures like triangle,rectangle and circle etc.The trained model helps user to draw figures using the data base labels.

I LOVE YOU

I NEED WATER

HELLO

DRAW A RECTANGLE

DRAW A CIRCLE

Fig. 2 Sample Images for hand gesture datasets

B. Image Acquisition and processing;

We crop our area of interest from the image captured.The region of interest form images of gestures have disturbance in background.To detect any foreground object to remove theses noises we calculate the weighted average for the

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11118

background and subtract this average from the frames that contain some object (hand in our case) as foreground. Then we find threshold value for every frame and determine the contours to return the max contours. Using contours we can identify if there is any foreground object.When any foreground object is detected then that frame instance is converted to black and white format, thresholded and save to the data base.

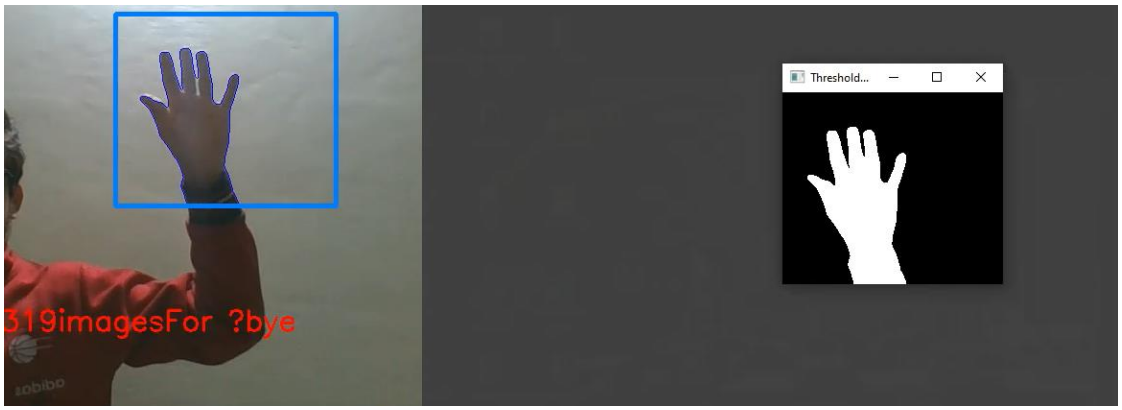

Fig. 3 ROI capturing and Thresholder Image

C. Convolutional Neural Network:

Convolutional neural networks are specifically designed to make inferences from visual data such as images and videos The features are extracted and learned to train the model, which gives better recognition accuracy compared to conventional Machine Learning algorithms. CNNs have numerous applications in the field of Signal Processing, robotics, medical imaging, data analysis, Business Intelligence, etc. The learning from unaltered and small data set with CNNs yields surprisingly better results.

CNN architecture is built with a combination of several layers that can be referred to as functional units of deep learning Convolution layers extract feature maps from the fed images. The feature maps are obtained by applying filters on the image termed convolution filters. The number of feature maps is equal to the number of filters. These filters are in the form of 3x3, 5x5, etc. matrices. The feature maps go through the activation function before feeding to the next layer. ReLU is one of the widely used activation functions.

D. Hard ware Components:

A good quality camera is required to capture the image. There are no specific requirements of camera but resolution of camera should be kept as low as possible .

E. Software Requirements:

- Tensorflow : To build and train the neural network
- Open CV : For image capturing and processing
- Numpy and Matplotlib : For mathematical tasks
- Tkinter : For creating an interactive UI.

IV. RESULTS

The model was trained over training data sets for 20,40 and 100 epoch cycles. The loss function was observed to decrease while an increase in accuracy of model was observed. The model accuracy grew in the range 92-96% for training and validation data. The increase in training and validation accuracy with each epoch indicate that the model is neither under fitting nor over fitting.The model reaches stability after 40 50 epochs.

The graphical representation of accuracy and is provided in the below. The increase in accuracy curve can be observed.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11118

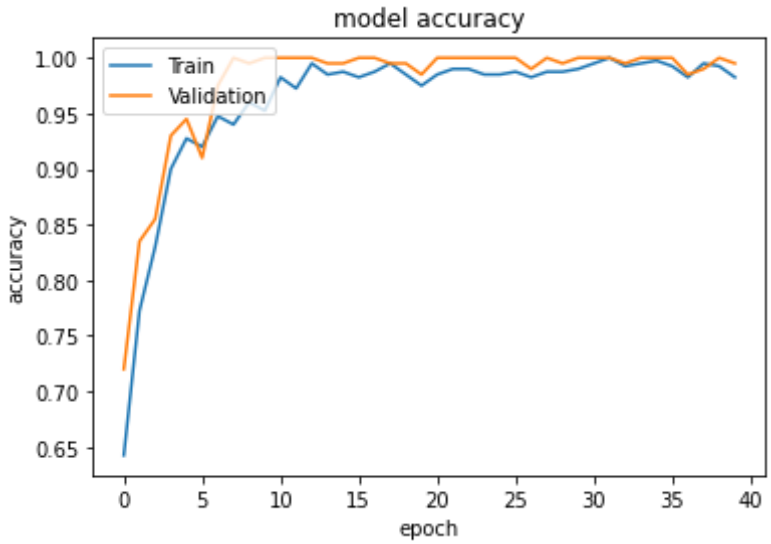

Fig. 4 Accuracy of model for datasets for 40 cycles

In certain instances, the model may not be able to differentiate between similar signs. This would be improved upon by further iterations and implementation of deep learning models to the algorithm.

V. CONCLUSION

Gesture recognition is a budding field of computer science and AI. Using hand gestures as the input to the system can enhance the way the user interacts with the system. This system can perform as a medium of communication for the deaf and dumb community of society. Thus tries to reduce the barrier for the after-mentioned minorities. The system also allows the user to add new gestures associated with labels to the system for translation.

The primary advantage of the system is that it is designed to be an interface that functions in real-time and would be available in masses. If developed as a mobile application then it can be effectively used by the targeted audience.

Like the general systems, this system does not promise 100% accuracy in translating the gestures. And since this is made to be used in real-time the chances of error increase due to randomness in the behaviour of the user and noise in the images captured. But as the user would be in the middle of a communication or task chance of the user to flag the error are considerably low. However, this does not rule out the probability of users ignoring all the errors. Inbuilt machine learning algorithms can be used to avoid errors flagged by the user.

Since the user has the privilege to add new gestures to the database for gesture translation. On which the model gets trained again. This process may affect the accuracy of the system. To overcome this, we have limited the number of new gestures that a user can add. But still, according to the nature of gesture, the model accuracy may get affected a bit.

Furthermore, this system does not translate the mood or emotions of the user. But it is made for simple translation of gestures to text.

Lastly, the system also has the feature of drawing geometric shapes according to the user's gesture. And is just one of the commands to demonstrate how gestures can be used as input to perform several tasks.

VI. FUTURE SCOPE

The system explains a very generic way of communication not restricted to any sign language. But while the users can add new gestures to the system the number of gestures that can be added is limited. The accuracy of CNN may vary on adding more gestures. Hence, algorithms that can build a flexible model that can adjust, modify and get trained with better accuracy on periodical addition of data could be used. And system with good computation power could be used to hold maximum gestures.

Sentiment analysis can be introduced to potentially increase the working of architecture.The system draws different geometric on getting commands from the user. Similarly, systems that can perform multiple tasks like playing music/video, sending email, playing a game, house automation etc. using gestures as input can be developed.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022
DOI: 10.17148/IJARCCE.2022.11118

REFERENCES

[1] Dr. V. Subedha, Sandhya , Shree Lakshmi , Swathi (IRJMETS - 2021) - Sign Language Recognition to Aid Physically Challenged Using Open CV and CNNN
[2] Shruti Chavan, Xinrui Yu and Jafar Saniie (IEEE - 2021) - Convolutional Neural Network Hand Gesture Recognition for American Sign Language
[3] Dr.J Rethna Virgil Jeny, A Anjana, Karnati Monica, Thandu Sumanth ,A Mamatha (IEEE -2021) - Hand Gesture Recognition for Sign Language Using Convolutional Neural Network
[4] Shagun Gupta ,Riya Thakur,, Vinay Maheshwari & Namita Pulgam (IEEE - 2020) - Sign Language Converter Using Hand Gestures
[5] Manasi Agrawal, Rutuja Ainapur, Shrushti Agrawal, Simran Bhosale,Dr. Sharmishta Desai (IEEE - 2020) - Models for Hand Gesture Recognition using Deep Learning
[6] Aishwarya Sharma, Dr. Siba Panda, Prof. Saurav Verma (IEEE -2020) - Sign Language to Speech Translation
[7] Ritika Bharti, Sarthak Yadav, Sourav Gupta, and Rajitha Bakthula (SSRN -2019) - Automated Speech to Sign language Conversion using Google API and NLP

