International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

Accord Computer through Hand Gestures using Deep Learning and Image Processing

Mr. Kasibhatla VSR Gautam Sri Charan[1], Mr. Sabbella Jnaneswar Reddy[2], Ms. S. Kavishree[3]

[1]Student, Dept. Of CSE, SCSVMV (Deemed to be University), Kanchipuram, TamilNadu, India
[2]Student, Dept. Of CSE, SCSVMV (Deemed to be University), Kanchipuram, TamilNadu, India
[3]Assistant Professor, Dept. Of CSE, SCSVMV (Deemed to be University), Kanchipuram, TamilNadu, India

---

Abstract - *Hand gesture recognition is one of the latest developments regarding Human-Machine interface. With keyboard and mouse already present to instruct and sensors using AI, this process of recognition of image will be much more effective. This technology would allow the operation of complex machines using only a series of finger hand movement. In our system, we will develop a behavioral model of Human-Machine Interface without any physical contact between them. Recently, gesture recognition has been a new developmental and experimental thing for most of the human related electronics. Also, this allows people to operate applications more conveniently. In our approach, we use deep learning algorithm to train the model and then classify various hand gestures. If the desired gesture is recognized, the respective command to the system will be triggered. Thereby controlling the system without any physical contact in a cost effective way. For this to achieve, we are using ResNet Architecture, a prominent deep learning algorithm, to train the model accordingly and recognize the gestures.*

Keywords: Deep Learning; Hand Gesture Recognition; Human-Machine Interface; Resnet Model;

1. INTRODUCTION

Humans have hands as a part of the body to communicate and interact with anything. The means of communication have been changing from time to time but there is a part of the process commonly being used ever since and that is hand. Keyboard is an important input device for a computer. Usually we type the keys in the keyboard which instructs the computer to enact accordingly. While hands playing this part, there are hand gestures which can do the same as they does in physical man to man communication. But, if a computer have the ability to understand the gesture and enact accordingly as a keyboard does, it would become a step up in the Human-Machine Interface(HMI). This can be achieved in many ways with the revolutionary growth in technology. Considering the rich quality of images that a basic web cam can provide, we can make it look so simple with the setup part. A hand gesture is always a variable means of communication that provides variable meanings. If a computer can recognize the feature of a gesture and classify from other gestures, we can command the computer to act accordingly. Some HMIs like robots, virtual objects are already playing major role in the world. There are both static and dynamic forms in presenting a gesture which facilitate and implement a successful interaction between the human and computer. In the view, there are several devices coming into existence like sensors, tracking devices, VR devices, etc. Even though revolutionary advancements taking place in technology, natural user interface(hand gestures) stands out as an effective and sophisticated means to deliver the desired output. There are many daily applications like Microsoft Powerpoint, Windows Media Player, which can be made better with such interface.

1.1 Objective

The main objective of this project is to implement communication between humans and computer without any physical contact. A computer is such a complex machine which needs input devices to command it. Nevertheless, there will be still some situations where the keyboard alone is not sufficient. In such cases, the hand gestures can provide an alternative, especially whatever idea we want to implement. To be precise, the objective is to

- Recognize the gesture performed by our hand using webcam.
- Extract the required feature and classify among others.
- Command the computer to do the operation.

1.2 Scope of the Project

The future scope lies in making the algorithms applicable for different gestures and there by different classification schemas can be applied. Also this project can further be developed by adding some features like

- Multi Gestures – It would be much more useful if we can use two or more motion that can be detected as a gesture for feature extraction. This would be an added

advantage as wide range of gestures are possible when multiple gestures are taken.
- Two handed gestures – It would be more convincible if two hands can be taken as a single gesture for feature extraction.

Also with ResNet having upgrading versions where the network connections in Architecture will be designed with much more ease, they can be implemented which can give outcomes with better accuracy and robustness.

2. PROPOSED SYSTEM

With this opportunity, we proposed a system which can fulfill the drawbacks of the existing system. Our system not only deals with dynamic gestures recognition but also cost effective. Our system doesn't need any physical contact to communicate with the computer. Also we don't need any gesture recognition devices. Just a webcam is enough. We will train the datasets consisting gesture video frames and build the model. We are using 3D Resnet 101 Architecture to train the model. In order to extract the features and recognize the gesture, the following steps are taken in our proposed method;

1. The data is prepared for training by creating dataframes from the downloaded dataset.
2. Then the prepared data is sent for model is trained and built.
3. Now a GUI is run which allows the user to capture the scene which is technically called Acquisition.
4. After capturing the video, the hand is detected in the frame and the gesture is analyzed.
5. Now the gesture undergoes feature extraction which takes place in ResNet itself.
6. Later, the gesture undergoes classification and finally the gesture will be recognized.
7. The recognized gesture is linked to the key task and therefore performs the operation.

2.1 Workflow

Our Project consists of very simple steps.

Step 1: Downloading dataset as this is deep learning and we require data to train the model.

Step 2: Then, we prepare our data from Pandas Library which is the coding library. In data preparation we create our data frames with the help of Pandas Library by uploading our data from system to our dataframe.

Step 3: Then the prepared data will be moved to Training of model. This is the lengthy part of our project. We use ResNet 101 Architecture and the base model(CNN) of this part. The feature extraction takes place here.

Step 4: The trained model will be used in classification process and by adding camera to the file, adding pictures from camera, taking pictures from video frames. Then add trained model is added before inputting the frames from camera.

Step 5: The output is viewed as the assigned task will be performed according to the matched gesture.

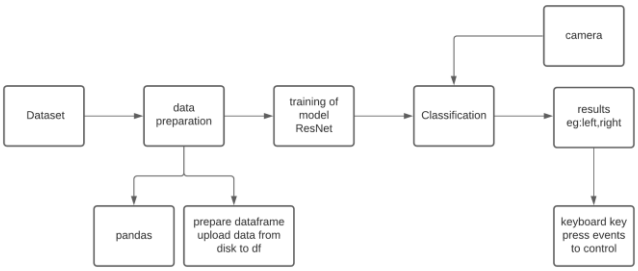


Fig-1. Project Workflow

2.2. Architecture – ResNet

Residual Networks, also called as ResNet, is a mere extension of the CNN Network. This network is mostly used in Computer Vision Projects. It had an ImageNet which is very beneficial. It allows to change very deep to 150+ layers and this feature is very useful. It can easily deal with plain network problems. The ResNet consists of a feature called skip connection, we can skip layers in between two or more layers. Theoretically, if we add more layers, we achieve higher performance and lower error. But in reality, with the skip connection that the ResNet is providing, the errors goes down and down. The skip connection convinces that instead of hoping every stack layers together, we directly jump layers. Simply apply skip layer and skip the one layer between that layers. The shortcuts connections are also used while jumping layers. So that is why, the ResNet can easily gain accuracy even there is a huge increase in depth. So we can apply different layers, skip connections, and can go deep into our objects and images and have a good trained model.

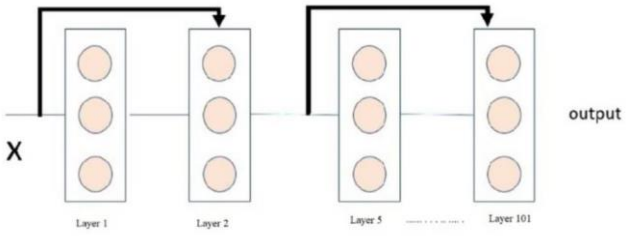


Fig-2. Residual Network – Skip Connection

2.3 Module Description

We will see as our project is divided into four modules which consists of all these steps and their preprocessing and internal processes and methodology.

2.3.1. Input Module:

This module involves in preparation of dataset. We already have the downloaded dataset which contains of frames of video in jpg format. Now we use Pandas Library for data preparation by creating data frames. We upload the downloaded data from our system to dataframe(df) using Pandas Library. The technique of data augmentation is used as a part of this module. Real time data augmentation, also known as online augmentation, is used as we are taking a huge dataset. This also consists of basic image processing techniques like flipping, cropping, colour channelling. We create batches from dataframes and process it for real time data augmentation.

```
In [10]: data = Dataloader(path_vid, path_labels, path_train, path_val)
         mkdirs(path_model, 0o755)
         mkdirs(os.path.join(path_model, "graphs"), 0o755)
```

```
        video_id                label
0          34870          Drumming Fingers
1          56557  Sliding Two Fingers Right
2         129112   Sliding Two Fingers Down
3          63861     Pulling Two Fingers In
4         131717     Sliding Two Fingers Up
...          ...                      ...
118557     75507             Swiping Down
118558     48433  Sliding Two Fingers Left
118559    146421  Sliding Two Fingers Right
118560     49514               Thumb Up
118561      4502     Sliding Two Fingers Up

[118562 rows x 2 columns]
       video_id                    label
0          9223                 Thumb Up
1        107090  Pushing Two Fingers Away
2         42920             Swiping Left
3        106485               Thumb Down
4        142201     Rolling Hand Backward
...         ...                      ...
14782     97044          Drumming Fingers
14783    136208  Sliding Two Fingers Right
14784     12180     Rolling Hand Backward
14785    119381               Thumb Down
14786     64033                Swiping Up

[14787 rows x 2 columns]
```

Fig-3. Loaded video labels to data using dataframes – INPUT MODULE

2.3.2. Training Module:

This module takes the prepared data and the model is trained using 3d ResNet 101 network. Normally a traditional CNN undergoes different layers to analyse the frames. Whereas, ResNet comes with an extraordinary feature called **skip** technology. Theoretically, the more the layers, the better the accuracy. But when applied, network becomes complex and so is time. This skip technology performs skips between layers so the network is lightened resulting in better accuracy. It also prepares a stack of layer as a function and directly skips to the function by using the concept of relu layer and pooling layer. Model with 16 frames has an accuracy of 93.71%, whereas model with 32 frames has an accuracy of 86.8%.

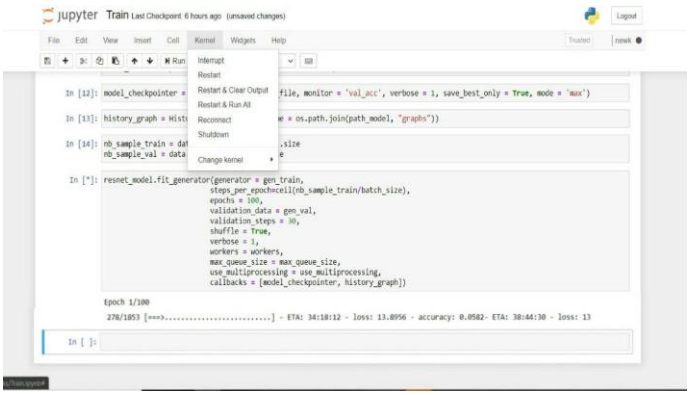


Fig-4. Training the model – TRAINING MODEL

2.3.3. Classification Module:

In this Module, the trained model is used. Now we add camera in this module which takes the pictures(frames) and processes in the trained model. We just have to feed the frames from the camera into the model. ResNet is such an architecture that the classification is done within with the input and recognize the gesture to act accordingly.

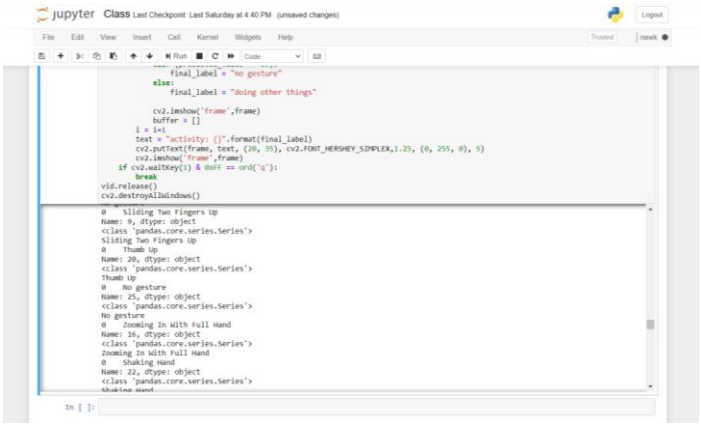


Fig-5. Gesture feature Extraction Successful – CLASSIFICATION MODULE

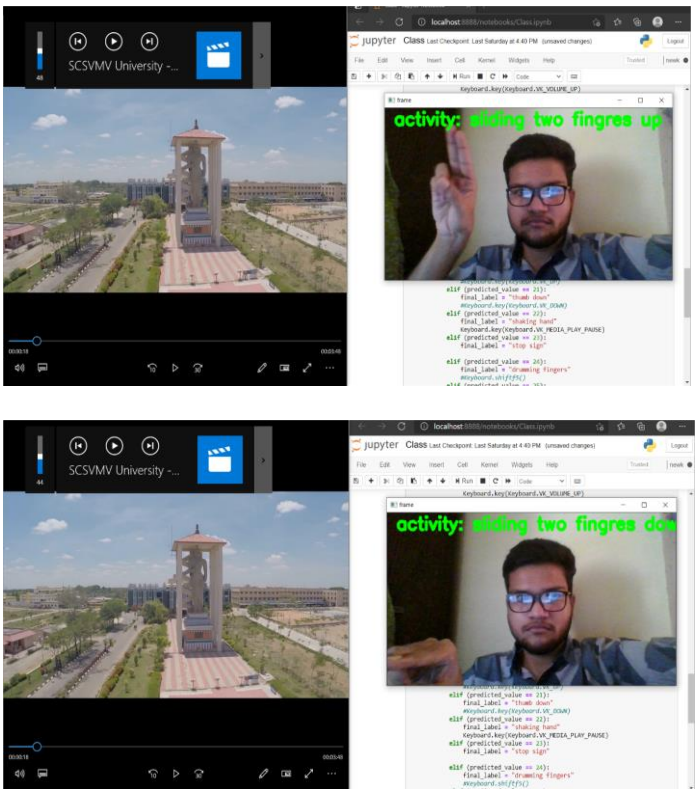
2.3.4. Output Module:

This module consists of GUI which consists of 400x400 size window that shows the video stream of what we do before the camera. This GUI provides us the information of what

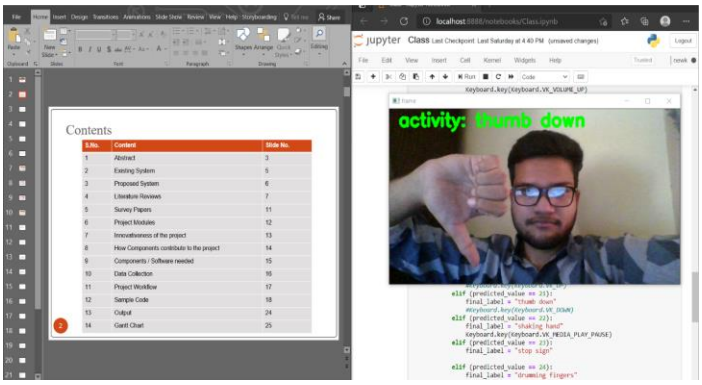
gesture we are making as the labels given for each type of gesture are taken from the dataframes.

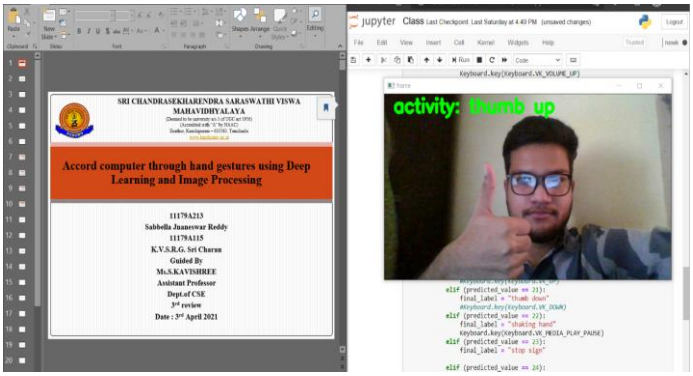
Fig-6,7,8. Successful Gesture Recognition – OUTPUT MODULE

3. RESULTS



Figs-9,10. Controlling Media Player using our system
(Sliding two fingers up – Volume up)
(Sliding two fingers down – volume down)





Figs-11,12. Controlling Powerpoint using our system
(Thumb down – Next slide)
(Thumb up – Previous slide)

4. VISUALIZATION

After training the model, the accuracy and the training loss with the epoch are plotted and saved in the directory mentioned. The trained model gives the statistics in a history graphs which are saved as .png file. The graphs obtained are showing that the resnet is giving higher accuracy and lesser loss than traditional CNN. The below are the results of the training data.

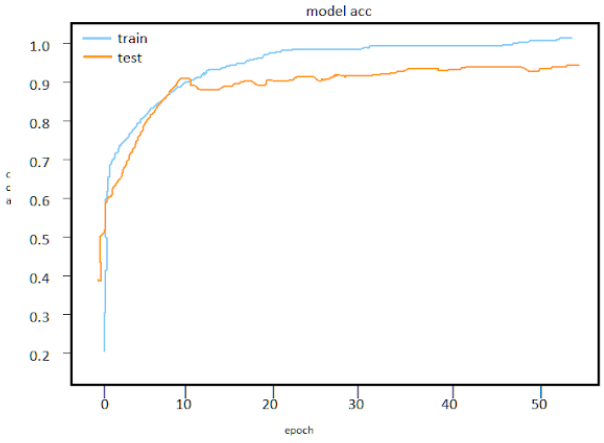


Fig-14. Train vs Test Model Accuracy

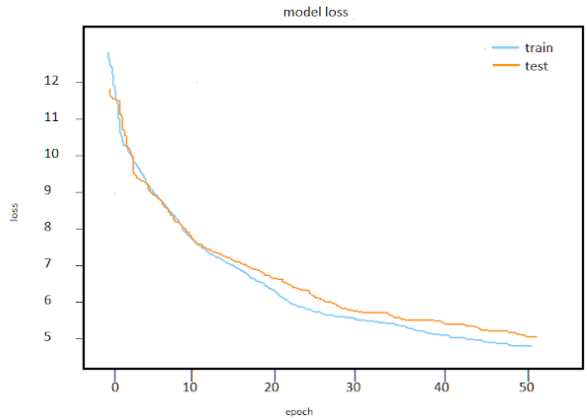


Fig-15. Train vs Test Model Loss

5. CONCLUSION

The goal of the project is to make user operate the computer without any physical contact. The implementation of natural interface to recognize hand gestures is a revolutionary advancement in the Human Machine Interface. Our project is solely brought into existence on the factors like sophistication and user satisfaction in a cost effective manner. And that is made possible with a contactless communication which is what we are implementing. We can just command a computer just with a slight movement of our hand.

REFERENCES

[1] Munir Oudah, Ali Al-Naji, Javaan Chahl. Hand Gesture Recognition Based on Computer Vision: A Review of Techniques: Electrical Engineering Technical College, Middle Technical University, Published on 23rd May 2020.

[2] H Pallab Jyoti Dutta, Debajit Sarma, M.K. Bhuyan and R. H. Laskar. Semantic Segmentation based Hand Gesture Recognition using Deep Neural Networks: Department of Electronics and Electrical Engineering, IIT Guwahati, Published on 25th June 2020.

[3] Dinh-Son Tran, NgocHuynhHo, Hyung-Jeong Yang , Eu-Tteum Baek , Soo-Hyung Kim and Gueesang Lee. Real-Time Hand Gesture Spotting and Recognition Using RGB-D Camera and 3D Convolutional Neural Network: School of Electronics and Computer Engineering, Chonnam National University, Published on 20th January 2020.

[4] Mohammad Sadegh Ebrahimi, Hossein Karkeh Abadi. Study of Residual Networks for Image Recognition: Stanford University, Published in April 2019.

[5] Andrés Jaramillo-Yánez , Marco E. Benalcázar and Elisa Mena-Maldonado. Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review: Artificial Intelligence and Computer Vision Research Lab, Department of Informatics and Computer Science, Published on 27th April 2020.

[6] Xianghan Wang, Jie Jiang, Yingmei Wei, Lai Kang. Research on Gesture Recognition Method Based on Computer Vision: Department of Systems Engineering, National University of defense Technology, Published in January 2018.

[7] Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun. Deep Residual Learning for Image Recognition: 2016 IEEE Conference, Las Vegas, Published on 12th December 2016.

[8] BAOQI LI, YUYAO HE. An Improved ResNet Based on the Adjustable Shortcut Connections: School of Marine Science and Technology, Northwestern Polytechnical University, Published on 23rd April 2018.

[9] Neha S. Rokade, Harsha R. Jadhav, Sabiha A. Pathan, Uma Annamalai. Controlling Multimedia Applications Using Hand Gesture Recognition: Student, Department Of Computer Engineering, GESRHSCOE, Published in August 2015.

BIOGRAPHIES

Kasibhatla VSR Gautam Sri Charan is pursuing Computer Science and Engineering in SCSVMV (Deemed to be University).

Sabbella Jnaneswar Reddy is pursuing Computer Science and Engineering in SCSVMV (Deemed to be University).

Ms. S. Kavishree is Assistant Professor in Computer science and Engineering department in SCSVMV (Deemed to be University).