ISSN No:-2456-2165

Hand Sign: An Incentive-Based on Object Recognition and Detection

Vikas and Rahul Mandal
Computer Science and Engineering,
Lovely Professional University, Punjab

Abstract:- The utilization of physical controllers like mouse, keyboards for HCI impedes the regular point of interaction as there is a solid boundary between the user and the PC. Hence, different strategies are assembled like speech, joint movement, and hand sign procedures to make it more natural and appealing. Over the most recent couple of years, hand gesture recognition has been viewed as an easy and normal procedure for human-machine communication. It is one of the methods for correspondence with PCs utilizing static and dynamic development and helps us perceive messages utilizing them. Numerous applications have been developed and upgraded for hand sign recognition. These applications range from cell phones to cutting-edge advanced robotics and from gaming to clinical science. In the vast commercial and research applications, recognition of hand signs has been performed by utilizing sensor-based wired installed gloves or by utilizing vision-based methods where skin tones, chemicals, or paperclips are utilized on the hand. In any case, it is attractive to have hand sign recognition techniques that are pertinent to a natural and bare hand. Today data of various researchers and now available to experiment with Hand Sign Recognition. We have used TensorFlow, OpenCV, and Jupyter Notebook for developing the Sign Recognition System where we have trained our model for various sign languages and alphabets. We have used Object Detection Technique to build this system where our webcam takes the input data and trains the system which is working in a virtual environment. Data accuracy depends on speed. So higher the speed lowers the accuracy and vice-versa. Using different hand signs to advance continuous application we pick a Vision-based Hand Gesture Recognition System that depends on various shape features.

Keywords:- *Human-Computer Interaction, Data Gloves, Optical Markers, Image-Based Technologies, Vision-Based Recognition System, OpenCV, Jupyter Notebook, Tensorflow.*

I. INTRODUCTION

Gesture-based communication is the method of correspondence that utilizes visual ways like expressions, hand movement, and body motion to convey meaning. It is a non-verbal mode of communication. This method is a boon for the deaf and dumb people as it can help to convey their message without any difficulties to any other person who don't have the knowledge of Sign Language. Gesture-based communication is incredibly useful for individuals who face trouble with hearing or speaking. Communication through sign language alludes to the change of hand motions into words or letters in order to communicate in dialects. In this way, the change of communication through sign language into words by a robust algorithm or a model which can help to overcome any issues between individuals with hearing or speaking disabilities with the rest of the world.

Computer vision and Machine Learning consist of a Vision-based Hand Sign Recognition System. As Sign convention is one of the easiest ways for human interaction, this region has numerous specialists that are chipping away at it, fully intent on making a Human-Computer Interaction (HCI) simpler and highly affordable. Thus, the essential objective of Sign recognition research is to make frameworks, which can recognize the human mode of communication and use them, for instance, to pass on data. For that, vision-based hand signal points of interaction require quick and incredibly vigorous hand identification, and motion gestures in real-time. Hand Signs are a strong human correspondence methodology with lots of possible applications and in this unique situation, we have gesture-based language recognition, the specialized strategy for deaf and dumb individuals.[1]

One of the primary objectives of Hand Sign Recognition is to make frameworks, which can recognize explicit signs and use them to pass data or to control a device. Apple Inc has developed this gesture control system in its ecosystem where users can swipe images from one Apple product to another. There are essentially two kinds of approaches for hand Gesture Recognition firstly the Vision-Based approaches and other via Electronic gloves which send data in form of electric signals and it is too expensive and complex architecture. Therefore Vison-Based framework is used as it is easy to access and manage. But it causes accuracy issues since it uses light as the mode to capture the image and if light intensity fluctuates the result can vary. Due to its broad domain of access, it can not only be used for disabled people but can be used in entertainment like gaming and animation, defenses, traffic, clinical domain, and much more. But communications via hand signs are not standardized and can cause misinterpretation.

II. MOTIVATION

According to the World Health Organization, One out of every four individuals counts to 2.5 billion people across the globe will suffer from mild-to-profound hearing loss by 2050. The criteria which WHO defines for disabling hearing loss is >40dB in adults and >30dB in children. According to the report, the cause of hearing loss is due to exposure to excessive noise, chronic ear infections, genetics, and aging.

ISSN No:-2456-2165

At present, 1.1 billion people aged between 12 to 35 years are at risk of hearing loss due to exposure to noise. It is really sad to say only 17% of those people use any hearing aid. It is due to the fact that a good hearing aid comes under a premium segment and also due to the lack of awareness among people. Prices of hearing aid can go from 50k to 3L and for most of the population in India, it is out of their pocket. Therefore, it is necessary to build a robust and affordable device that can reach millions of people who are in need of it. It can be achieved via ML and Object Detection techniques. That's why we took the initiative to work on it. [2]

III. LITERATURE REVIEW

Understanding the existing system plays a crucial part in any research paper. It provides a deeper knowledge and forms the foundation for existing system optimization. Therefore, we have conducted a literature review on Hand gestures and their existing algorithm from various sources.

- TensorFlow an AI framework developed by Google works on a general scale and in heterogeneous conditions. Its purpose has been dataflow diagrams to address calculation, shared state, and the tasks that change state. It has the power to map the hubs of a dataflow diagram across many machines in a bunch, and inside a machine across numerous computational gadgets, including multicore CPUs, universally useful GPUs, and hand-crafted ASICs known as Tensor Processing Units (TPUs). This engineering has given adaptability to the application designer: while in past "boundary server" plans the administration of the shared state is incorporated into the framework. It also empowers engineers to explore different avenues regarding novel enhancements and preparing calculations. TensorFlow backings an assortment of uses, with attention on preparing and derivation on profound brain organizations. [3]
- The point of picture handling has been to assist the PC with grasping the substance of a picture. OpenCV is a library of programming capacities essentially utilized for picture handling. Its point of interaction has been written in C++ to handle the processing of the system swiftly. It gives a true standard API for PC vision applications. It has the ability to tackle some constant issues utilizing picture handling applications. [4]
- Because of object identification's close relationship with video analysis and image understanding, it has drawn many researchers in recent years. Conventional article discovery techniques are based on high-quality highlights and shallow teachable structures. Their exhibition effectively deteriorates by building complex gatherings which consolidate various low-level picture highlights with significant level settings from object finders and scene classifiers. With the fast improvement in profound learning, all the more incredible assets, which has had the option to learn semantic, significant level, further highlights, has acquainted which address the issues existing in customary structures. These models act distinctively in network design, preparing technique and enhancement work, and so on. In this paper, they give a survey of profound learning by putting together image recognition frameworks. The center with respect to ordinary nonexclusive article location models alongside certain alterations and helpful stunts to further develop discovery execution. As particular explicit identification assignments show various qualities, they additionally momentarily study a few explicit undertakings, including remarkable article recognition, face location, and passerby discovery. Trial examinations are likewise given to look at different techniques and make a few significant determinations. At last, a few promising headings and assignments are given to act as rules for future work in both article identification and pertinent brain network-based learning frameworks. [5]
- In the journal "Appearance Based Recognition of American Sign Language using Gesture Segmentation" Kulkarni introduced the static posture of American Sign Language with the help of the Neural Network Algorithm. For 26 alphabetic characters, she used 8 samples 5 for training and 3 for testing in MATLAB which resulted in 92.78% accuracy. She used the histogram technique and Hough Algorithm to extract features of the HSV color image with uniform background.[6]
- In the journal "New Approach to Hand Tracking and Gesture Recognition by a New Feature Type and HMM" Pham et al introduced Vietnamese Sign Language. It consists of a vocabulary of 29 gestures. Here, the system consists of 3 modules: Real-time hand tracking, training gesture, and pseudo-2-D hidden Markov models. They used the Tower method and skin color to track the hand region. [7]

IV. IMPLEMENTATION OF THE PROPOSED WORK

A. Image Workflow and Data Examination:

This paper contains, automatic recognition of the finger spelling in the Indian sign Language. Here, the sign act as a input to the framework. Further different advances are performed on the input data of sign image. Right off the bat segmentation phase is performed based on the skin tone to detect the sign state. The detected area is then changed into binary picture. Afterward, the Euclidean distance transformation is applied on the binary image. Row and column projection is applied on the distance changed image alongside HU's minutes are utilized.[8]

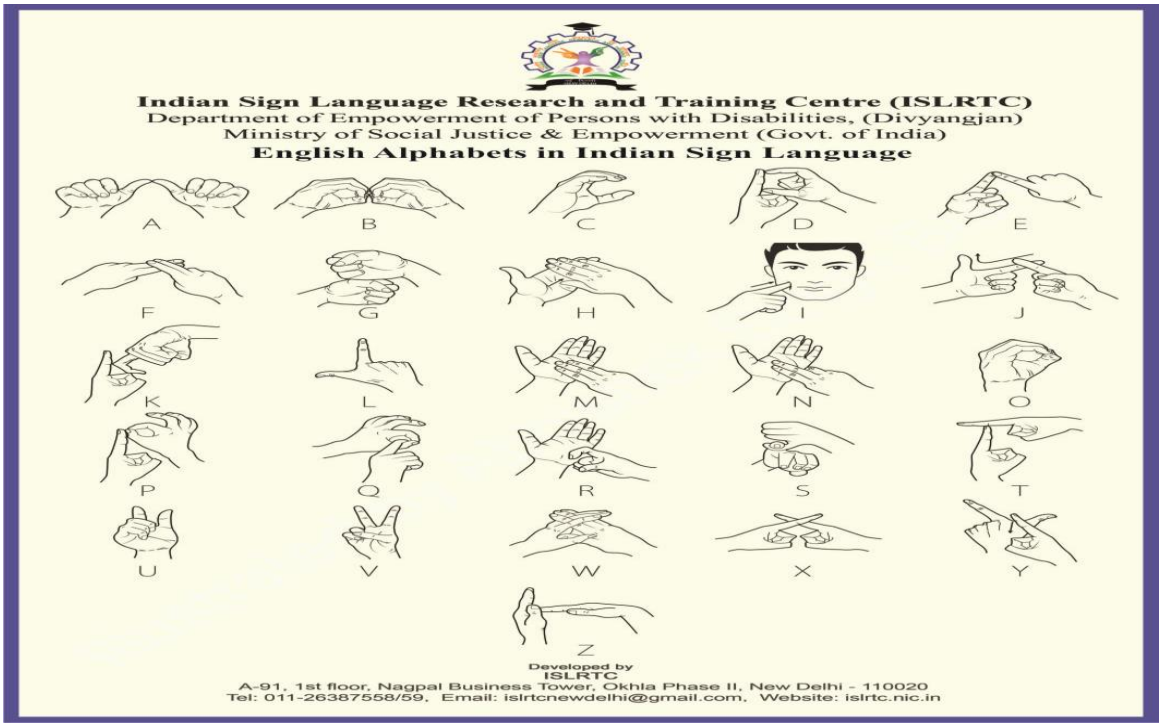

Fig. 1: English alphabet in Indian Sign Language Approved by Indian Sign Language Research and Training Center (ISLRTC)

B. Build an Input Pipeline

Tensorflow API enables us to build complex input pipelines from simple, reusable pieces.In our case, we have used the pipeline of an image model(SSD MobileNet V2 FPNLite 320x320 mode)[9] to collect the data from files in the operating system, apply the technique that adds 'noise' to a dataset, or in this case file to allow individual record confidentiality(data perturbations) to each image and merge selected images into a TFRecords format which is a binary file format for storing data for training. To record connects our image file and annotation file which we created with the help of a label map in our environment and help to train our model with more efficient storage, Fast input/output, and self-contained files. The pipeline for an image model involves a Feature Pyramid Network(FPN) [10]which is a feature extractor that takes a single-scale image of arbitrary size as input and outputs proportionally sized(In our case it is 320x320 resolution) feature maps at multiple levels, in a fully convolutional fashion. The image model is going to compress the image from 604x480 from the webcam to 320x320 in the preprocessing stage with the help of an image resizer and take the detection and revert back in the post-processing stage. Thisprocess is the backbone of the convolution architecture from the MobileNet V2[9][11].

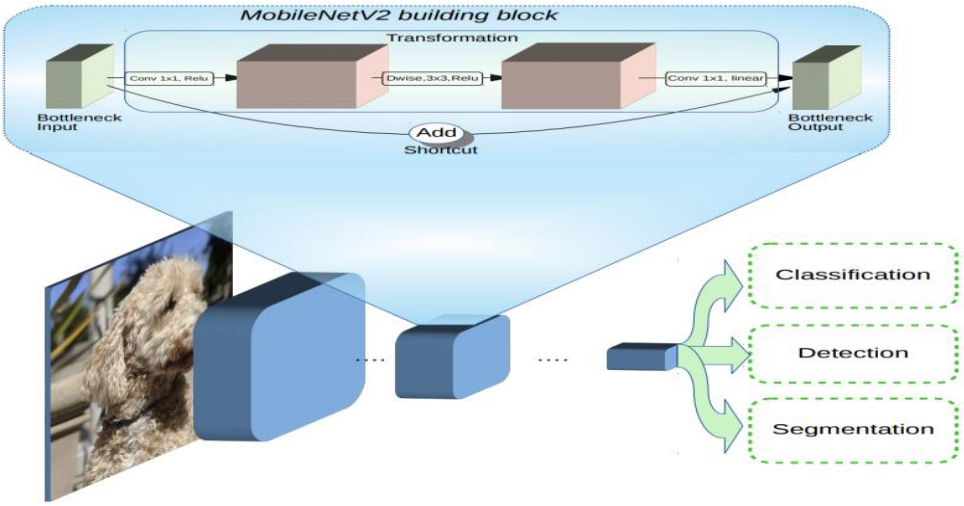

Fig. 2: MobileNet V2 Building Block

ISSN No:-2456-2165

C. Building the custom model from the pretrained model:

Given the basic pipeline with the model SSD MobileNet V2 FPNLite[12]320x320 model, we have to configure our paths and file for fine-tuning for example label map, training, testing, and the evaluate the input(testing) which we have created the with help of TFRecords and our subject(sign language) testing and training dataset.

D. Training and Testing of the Model:

Code Snippet:

Python Tensorflow\models\research\object_detection\model_main_tf2.py --model_dir=Tensorflow\workspace\models\my_ssd_mobnet --pipeline_config_path=Tensorflow\workspace\models\my_ssd_mobnet\pipeline.config --num_train_steps=2000

Now the configured pipeline file (pipeline.config) we have created according to our dataset from the above code we can initiate the model_main_tf2.py file from the TensorFlow object detection library to our pipeline.config file which we created with the ssd_mobilenet v2 fpv lite 320x320 architecture and specify the training steps which is using one batch size of training data to train the model and repeat the process 2000 times.

E. Model Improvement:

Now after the initial steps, we are going to do the performance tuning to improve the model by:

- Adding more images of the low-performing classes to the training set.
- Training for longer by increasing the training steps.
- By changing the architecture in this case we have used SSD MobileNet V2 FPNLite 320x320. We can change it according to our needs in speed(ms) and accuracy(mAP).

V. EXPERIMENT DESIGN AND RESULT

Use of SSD MobileNet V2 FPNLite 320x320 model.

This is a pre-trained model on the COCO 2017 dataset(COCO is large-scale object detection, segmentation, and captioning dataset)

This model can be used on various environments such as the cloud platform(google-colab), various operating systems, raspberry pi, web technologies like ReactJs with the help of TensoflowJs, and mobiles due to being its Feature Pyramid Network Lite pack.

We have also compared the SSDLite(Single Shot Detection) with the FPNLite(Feature Pyramid Network) and find FPNLite is much better for small object detection. It was due to the better feature extraction which selects the right set of features in the decisive key in order to avoid ambiguity. The feature extractor is more easily differentiated from any pattern. In the case of Sign language, it's necessary to choose features from different angles, rotation, and inversion to extract the image and decomposed it into components with the help of depthwise convolution for the filtration. For sign language, it uses the basic shape boundary information with the complete ignorance of other details and due to the data augmentation, it gives us a more accuracy rate.

VI. CONCLUSION AND FUTURE WORK

We have presented a clean and simple virtual platform to build the sign detection model and implement it on the web for its environment uses. Our method shows a significant improvement in our model improvement steps with various technologies such as the Tensorflow Object Detection module and COCO dataset trained model in an anaconda Jupyter notebook with the help of an open vision detection library.

We have come a long way in the Sign Recognition Detection but it is necessary to build a more robust and affordable product in the segment of the Sign Detection so that it can be commercially available to everyone who needs it. Its domain is vast and integrating it with the various devices requires a lot of work. We can work with cloud computing so that Sign Detection can be used in low-end devices which takes a lot of processing time. It is also important to standardize sign language so that most people around the globe can utilize it. But standardizing is not easy as there are various countries with different languages which can cause a lot of misinterpretation. We can integrate it with various IoT devices to make it portable and can help disabled people who need them the most.

REFERENCES

[1.] S. R. ReddyGari, R. Rumana and R. Prema, "A Review Paper on Sign Language Recognition for The Deaf and Dumb," *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY,* vol. 10, no. 10, 2021.

[2.] "DowntoEarth," 02 03 2021. [Online]. Available: https://www.downtoearth.org.in/news/health/every-4th-person-to-suffer-hearing-loss-by-2050-who-75718.

[3.] M. Abadi, . P. Barham, J. Chen, P. Warden, G. Irving, X. Zheng and Y. Yu, "TensorFlow: A system for large-scale machine learning," *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), USENIX Association,* pp. 265-283, 2016.

[4.] N. K. Mahamkali and V. Ayyasamy, "OpenCV for Computer Vision Applications," in *Proceedings of National Conference on Big Data and Cloud Computing,* Trichy, 2015.

[5.] Z. Q. Zhao, P. Zheng, S. T. Xu and X. Wu, "Object Detection with Deep Learning: A Review," *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS,* pp. 1-21, 2019.

[6.] V. S. Kulkarni and S. D. Lokhande, "Appearance Based Recognition of American Sign Language using Gesture Segmentation," *International Journal on Computer Science and Engineering (IJCSE),* vol. 2(3), pp. 560-565, 2010.

[7.] P. TheBao, N. ThanhBinh and T. Khoa, "A New Approach to Hand Tracking and Gesture Recognition by a New Feature Type and HMM," *Fuzzy Systems and Knowledge Discovery, Sixth International Conference,* vol. 4, pp. 3,6,14-16, 2009.

[8.] P. Jadav and Y. I. Rokade, "Indian Sign Language Recognition System," *International Journal of Engineering and Technology*, pp. 189-196, 2017.
[9.] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Computer Vision Foundation.*
[10.] T. Y. Lin, P. Dollar´, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," in *Facebook AI Research (FAIR),Cornell University and Cornell Tech*, 2017.
[11.] "Tensorflow," [Online]. Available: https://www.tensorflow.org/guide/data.
[12.] G. Ghaisi, T. -Y. Lin, R. Pang and Q. V. Le, "NAS-FPN: Learning Scalable Feature Pyramid Architecture," 2019.