International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

Hand Gesture Recognition using YOLOv5

Nikhil Jadhav[1], Vrushabh Kakkad[2], Vaibhav Patil[3], Sanket Vetal[4]

[1-4]*Department of Information Technology, JSPM'S Rajarshi Shahu College of Engineering, Pune, Maharashtra, India*

Abstract - *Hand Gesture Recognition which is a part of language technology involves evaluating Human Hand Gestures through Machine Learning Algorithms. People use this technology to interact with computers or electronic devices without touching them. Computer vision algorithms or cameras are used to convert these sign languages. This is one of the primitive connectors between man and machine. This makes it one of the important applications of Graphic User Interface (GUI). The process involves detection of Hand portion and then finding out the gestures through sign language made by hand. Major application of this technology is in the automotive sector.*

Key Words: Machine Learning algorithm, Computer vision, sign language, Graphic User Interface, automotive, detection.

1. INTRODUCTION

Hand Gestures are an aspect of body language that can be conveyed through finger position and shape constructed through palm. First, the hand region is detected from the original images from the input devices. Then, some kinds of features are extracted to describe hand gestures. Last, the recognition of hand gestures is accomplished by measuring the similarity of the feature data. Any random hand gesture consists of four elements as hand configuration, movement, orientation and location. These gestures are further classified as static gestures and dynamic gestures. The proposed system deals with static gestures. Hand gesture recognition has great value in many applications such as sign language recognition, augmented reality (virtual reality), sign language interpreters for the disabled, and robot control.

"YOLO" is an acronym for "You Only Look Once". It is the state-of-the-art object detection framework. The Yolov5 algorithm covers mainly these steps: Image Classification, Object Localization and Object Detection. There are multiple variants in YOLO depending on the size and speed of the model. YOLOv5s is the smallest yet fastest variant. YOLOv5m is a medium sized model which is slower than YOLOv5s but faster than other variants.

The traditional method uses sensors. These sensors detect a physical response according to hand movement or finger bending. The data collected were then processed using a computer connected to the glove with wires.

The proposed system is implemented using the YOLOv5l model. It is a robust object detection framework. This system requires only a camera and a computer for detection. The proposed system has precision of 0.7, with 0.95 recall value. This system has mAP@0.5:0.95 is 0.77.

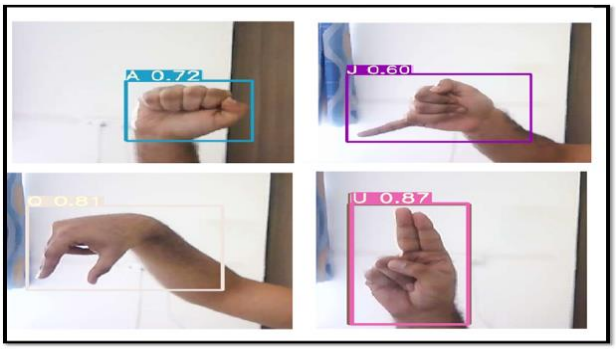

Fig - 1: Sign Language Letters Recognized

2. WORKING

The traditional system uses specially designed hand gloves and sensors. The sensor recognizes the position of fingers, its alignment and transfers it to the system for sign language detection. Due to excessive dependency on sensors, any kind of damage to it can lead to complete failure of the system. The sensor needs to be calibrated before use as per end-user palm physique. This contributes to the increase in initial setup, maintenance and hardware cost.

For the proposed system, a YOLOv5l (large) model is used which is larger in size and highly accurate. Smaller models can also be used for portable systems with slight compromise in accuracy. In the proposed system, images are captured using a camera and sent to the back-end. In the back-end, this captured image is passed into the YOLOv5 model, which outputs and displays that image with a bounding box around the detected palm area. This also displays the letter made using sign language, above the bounding box. This process is continued in a loop until the stop button is pressed.

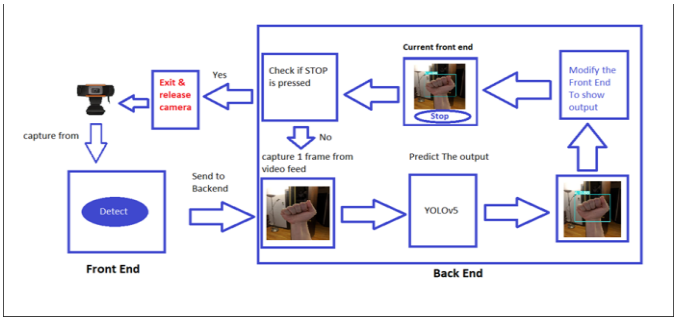

Fig -2: System Architecture for Proposed System

If needed, a portable system, YOLOv5s model can be trained on a Raspberry Pi computer, which is a very small yet powerful computer. For increased scalability, the application can be deployed on a server and then access the application by just using browser and webcam. This also helps in reducing hardware expenses. Because of deploying the application on the server, the only required hardware is the server itself. And the client can access this application using any browser, without purchasing extra hardware.

Desktop applications can also be operated using YOLOv5x (extra-large) model for uncompromised accuracy and performance.

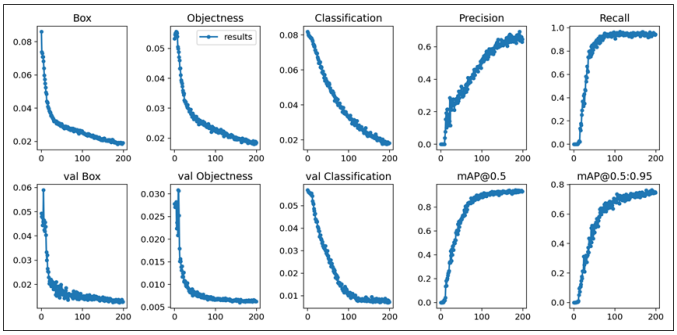

Fig - 3: Observation Graph

3. ADVANTAGES

1) Functioning:

1.1) The output of the sign language will be displayed in the text form in real time.

1.2) The communication with hearing and speech impaired people become easier.

1.3) The images captured from a continuous video stream of a webcam are passed in a yolo model and the output is text. Thus, this feature of the system makes communication very simple and delay free.

2) No external hardware required:

2.1) As traditional approach is using sensors or gloves, the implemented system suggests the direct use of images from a web camera for recognition which reduces the cost.

2.2) The only requirement here is a webcam, a computer (for dedicated application) or a server (for a web application which is to be deployed on a server).

3) Portable:

3.1) If the Entire project is implemented on a Raspberry Pi computer, the entire system becomes portable and can be taken anywhere.

3.2) Deploying the project on a web application will increase its access without having to setup manually.

3.3) Deploying this project as a dedicated Desktop application will enable the model to use a dedicated GPU on it which in turn will increase its performance.

4) Does not get damaged through use:

4.1) As no specific hardware is required here, and the project is just a software, there is no degrade in performance with respect to time.

4.2) The only variable factors for the proposed system are that it depends on how powerful the computer it is deployed on is.

4. DISADVANTAGES

1. The proposed system uses only cameras for input purpose, so illumination of the room shadows, background can hinder the output.
2. Palm must be facing towards the camera for accurate results.
3. The performance of the system also depends on the type of hardware used. Higher end systems can yield higher accuracy as well as high frames per second.

5. CONCLUSIONS

The traditional approach seems to be quite expensive and a complex hardware system is required. This also demands high maintenance. On the other hand, the proposed system focuses on recognizing gestures through the camera, thus reducing hardware requirement, complexity and is cost effective. Both techniques mentioned above, however, have their advantages and disadvantages and may perform well in some challenges while being inferior in others.

ACKNOWLEDGEMENT

The authors would like to thank the project guide Prof. Nidhi Ma'am for this opportunity, valuable support and encouragement.

REFERENCES

[1] https://en.wikipedia.org/wiki/Gesture_recognition

[2] Gesture Recognition Developments- Vrushabh Jayesh Kakkad, Volume:07 Issue: 12, Dec 2020, e-ISSN: 2395-0056, p-ISSN: 2395-0072.
https://www.irjet.net/archives/V7/i12/IRJET-V7I12365.pdf

[3] Hand Gesture Recognition Based on Computer Vision: A Review of Techniques Munir Oudah 1, Ali Al-Naji 1,2, * and Javaan Chahl 2, published: 23 July 2020
https://doi.org/10.3390/jimaging6080073

[4] https://www.ijser.org/paper/Sign-Language-Recognition-System.html#:~:text=The%20output%20of%20the%20sign,displayed%20at%20the%20same%20time.

[5] https://github.com/ultralytics/yolov5

[6] https://www.hindawi.com/journals/tswj/2014/267872/