International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Real Time Assistive System for Deaf and Dumb Community

Ashish Dandade, Pranav Tayade, Rushikesh Patil, Jayant Mitkari

[1]Ashish Dandade, Information Technology , SSGMCE Shegaon.
[2]Pranav Tayade, Information Technology , SSGMCE Shegaon.
[3]Rushikesh Patil, Information Technology , SSGMCE Shegaon.
[4]Jayant Mitkari, Information Technology , SSGMCE Shegaon.

Abstract - People communicate with other people by utilizing words that come from various different tongues, but if somebody is deaf or dumb and incapable to listen or talk? People who are deaf or dumb must use a particular sign for communication purposes. A person with translation skills is required since not everybody grasps sign language, which breaks down interaction between sensing of deaf/dumb people. An automated model that identifies hand gestures and conduct the same job as translators could have a beneficial impact on the social lives of these hearing and speech impaired people because interpreters aren't always readily available. Applying the YOLOv5 machine learning model, hand motions have been learned and predicted. It will have many applications such as, online communications, Touchless operating, Human-Computer Interaction (HCI).

Key Words: Machine Learning, YOLOv5, Sign Language, Python.

1.INTRODUCTION

63 million deaf and dumb people live in India, where there are 2.4 million of them. These individual lacks the conveniences that a typical person would have. The main cause of this is a lack of communication since stupid people cannot talk and deaf people cannot listen. Fig.1 displays an analysis of a survey [1].Sign language is used by those who have difficulty in hearing or speaking as a means of communication. In sign language, people express their thoughts and feelings via nonverbal gestures. However, non-signers find it very difficult to understand, thus skilled sign language interpreters are needed during training sessions as well as sessions for legal, medical, and educational uses.

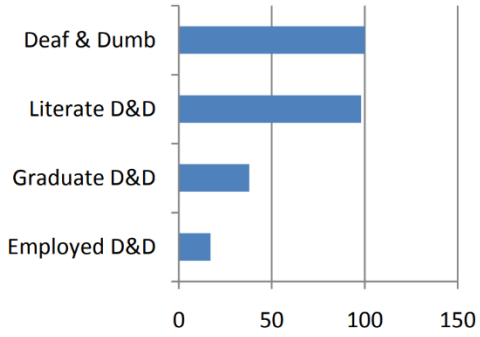

Fig. 1: Deaf and Dumb Work Survey

Over the preceding five years, there has been an increase in demand for translation services. More techniques are now readily available, including high-speed video remote human interpretation. As a result, they will provide a service that is easy to use but has significant disadvantages for sign language interpretation. In order to solve this, we use a combination of two models to differentiate between sign language motions. Using video footage from the distinct Sign Language Dataset, we train the computer to detect motions. A range of motions that were performed repeatedly in various video and context settings are included in the collection.

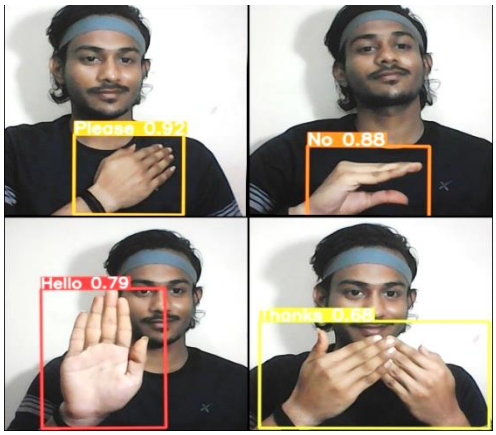

Fig. 2: Sign Language Word Recognized.

For convenience, the videos have been captured at a consistent frame rate. In contrast to the more traditional "sliding window" technique, which frequently requires numerous classifications runs for particular picture segments, our "You Only Look Once" approach analyses input images whole. This method speeds up processing since fewer separate assessments are required, and accuracy is increased because the whole neural network has access to the global visual context.

2. LITERATURE REVIEW

A. Related Work

To recognize the gesture, a lot of effort has already been done. Two sorts of techniques are now in use. One method makes use of a specific device to detect hands and identify the gestures they represent, while the other makes use of deep learning.

1) Device-based identification: For the Kinect sensor [2], the proposed technique first detects the segments of signs using hand movements before verifying the handshapes of the segmented signs using a Boost Map embedding method. It gathers 3D depth data about the hand movements using a Microsoft Kinect sensor. [3] uses a sensor glove to record the gesture, which is subsequently recognized and classified using an Artificial Neural Network (ANN). [4] uses the flex sensors

to capture the movement of the fingers and movements of the hands before capturing the posture and tries to identify if the posture falls into any pre-existing categories. A suitable matching method is employed to identify the precise value of the posture based on the belonging category. The Vietnamese language is recognized using this methodology. Using Microsoft Kinect, Convolutional Neural Networks (CNN), and GPU acceleration, a recognition system was suggested in [5]. The algorithm has a high degree of accuracy in identifying 20 Italian gestures.

Typically, device-based identification operates over two phases. The device initially collects the sensory signal from the gesture before utilizing that signal to identify the sign in the subsequent step. This method usually yields accurate information, but it is quite cumbersome for daily use, and the equipment is still expensive to run.

2) Deep learning-based identification: In computer vision and deep learning-based recognition, the useful feature maps from an input picture are obtained using a neural network with hidden layers. Many deep learning networks, including AlexNet [6], VGGNet, ResNet, and others, do exceptionally well in object detection. Additionally, there are several region-based models like R-CNN [7], Fast R-CNN [8], and Faster R-CNN that attempt to locate the area of an image where there is a likelihood of finding the item before attempting to ascertain if the object is really there in that area. Two-stage object detectors are the name given to these detection techniques. There are several ways to recognize hand motions using these cutting-edge CNN models. [9] KNN and SVM are used to assess the classification after using various image processing techniques. Convert ASL into the English alphabet using a pre-trained Google Neural Network architecture.

B. YOLO

Convolutional Neural Network (CNN) technology was used in the development of YOLO [2], which produces quick and accurate object recognition. By just seeing the input photos once through the neural network, the YOLO (You just Look Once) technique predicts the identified item in the image. It operates by segmenting the input picture into several grids depending on predetermined grid sizes, and then it forecasts the likelihood that the desired object will appear in each grid. In a single run of the algorithm, it predicts every class and object bounds that are present in the picture. As a result, it developed into an extremely quick end-to-end object detection that is also suitable for real-time detection. The YOLO algorithm has also been continuously improved [6] with regard to accuracy, speed, and lightweight.

3. METHODOLOGY

The conventional approach makes use of sensors and hand gloves that are particularly made. The sensor provides the finger position and alignment information to the system for sign language recognition. The system is overly dependent on the sensors, therefore any damage to them might result in system failure. Prior to usage, the sensor must be calibrated to the end-user's palm anatomy. The cost of initial setup, ongoing maintenance, and hardware is increased as a result of this.
The YOLOv5 (large) model, which is bigger and more accurate, is utilized for the suggested system. For portable systems, smaller versions can be employed with a modest accuracy penalty. Images are taken using a camera and communicated to the back-end in the proposed system. The YOLOv5 model receives this taken picture in the back-end and produces and displays it with a bounding box around the recognized palm region. Above the bounding box, this also shows the sign language letter. It repeats this procedure until the sign changes. Shown in fig. 3

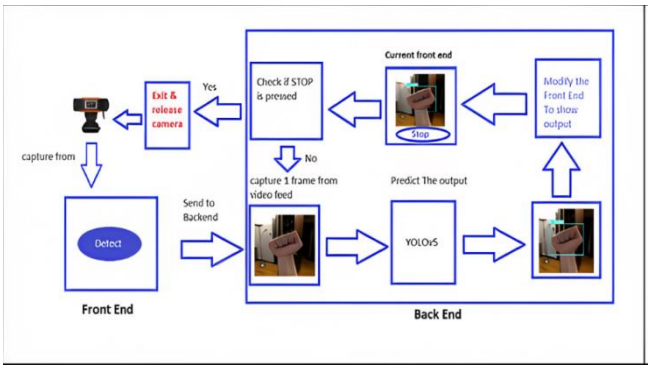

Fig.3 System Architecture for Proposed System

If a portable system is required, the Raspberry Pi computer, which is a tiny but potent computer, can be used to train the YOLOv5s model. The programs may be installed on a server for greater scalability, and users can then access it by just using their browser and webcam. This lowers the cost of the hardware as well. The server is the sole piece of hardware needed because the application is deployed on the server. Additionally, the customer doesn't need to buy any additional hardware to use any browser to access this programs. The YOLOv5x (extra-large) variant may be used to control desktop programs with uncompromising precision and performance.

4. ADVANTAGES

1) Performance:

1.1) The sign language output will be shown in text form in real time.

1.2) It becomes simpler to communicate with those who have hearing or speech impairments.

1.3) A yolo model is used to process the pictures from a webcam's continuous video stream, and the result is text. As a

result, this system feature makes communication incredibly easy and delay-free.

2) No additional hardware is needed:

2.1) The implemented method recommends the direct use of pictures from a web camera for recognition, which lowers the cost, as opposed to the usual way of employing sensors or gloves.

2.2) In this case, a webcam, a computer (for a dedicated program), or a server (for a web application that will be installed on a server) are the only prerequisites.

3) Transportable

3.1) The system as a whole becomes transportable and portable if the complete project is run on a Raspberry Pi computer.

3.2) The project's accessibility will be increased without the need for manual setup by being deployed on a web application.

3.3) The model's performance will be improved by using a dedicated GPU when this project is launched as a standalone desktop program.

4) Does not suffer harm from use:

4.1) Because this project just requires software and no special hardware, there is no performance degradation with the system.

5. DISADVANTAGES

1. The suggested system solely accepts input from camera since room shadows and backgrounds might interfere with the output.
2. For precise results, the palm must face the camera.
3. The kind of hardware being utilized affects the system's performance as well. better end systems may provide high frames per second and better precision.

6. CONCLUSIONS

The conventional method appears to be fairly costly, and a sophisticated hardware setup is needed. This necessitates a lot of upkeep. The suggested method, on the other hand, concentrates on recognizing motions through the camera, which lowers the amount of hardware needed, reduces complexity, and is more affordable. However, each of the mentioned methods has benefits and limitations, and they may be superior in some situations while performing well in others.

ACKNOWLEDGEMENT

The authors would like to thank the project guide Prof. Mrs. P.V. Kale for this opportunity, valuable support and encouragement.

REFERENCES

1. Laura Dipietro, Angelo M. Sabatini and Paolo Dario ―A Survey of Glove-Based Systems and Their Applications‖, IEEE Transactions on Systems, Man and Cybernetics—Part C: Applications and Review, Vol. 38, No. 4, pp. 461-482, July 2008.
2. S. A. Mehdi and Y. N. Khan, “Sign language recognition using sensor gloves,” in Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP’02., vol. 5. IEEE, 2002, pp. 2204–2206.
3. L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, “Sign language recognition using convolutional neural networks,” in European Conference on Computer Vision. Springer, 2014, pp. 572–578.
4. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, vol. 25, pp. 1097–1105, 2012.
5. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning. thirty-first aaai conf,” Artif. Intell, 2017.
6. R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.
7. S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards realtime object detection with region proposal networks,” arXiv preprint arXiv:1506.01497, 2015
8. R. Sharma, Y. Nemani, S. Kumar, L. Kane, and P. Khanna, “Recognition of single handed sign language gestures using contour tracing descriptor,” in Proceedings of the world congress on engineering, vol. 2, 2013, pp. 3–5.
9. Geethu G Nath and Arun C S, "Real Time Sign Language Interpreter," 2017 International Conference on Electrical, Instrumentation, and Communication Engineering (ICEICE2017).
10. Kumud Tripathi, Neha Baranwal and G. C. Nandi, "Continuous Indian Sign Language Gesture Recognition and Sentence Formation", Eleventh International MultiConference on Information Processing-2015 (IMCIP-2015), Procedia Computer Science 54 (2015) 523 – 531.
11. Manasa Srinivasa H S and Suresha H S, "Implementation of Real Time Hand Gesture Recognition," International Journal of Innovative