JSRT JOURNAL

PAGES: 188-192
9/10/23
JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY(JSRT) VOLUME-1 ISSUE-6 SEPTEMBER
Registered under MSME Government of India ISSN: 2583-8660

YOLOv5 SignSense: Empowering Deaf and Mute Communication through Gesture Recognition

Summaiya Yasmeen[1], Dr. Shameem Akther[2]

[1]*M. Tech Student, Computer Science and Engineering department, KBN University, Kalaburagi, India, summaiya6398@gmail.com*

[2]*Assistant Professor, Computer Science and Engineering department, KBN University, Kalaburagi, India, shameem@kbn.university*

ABSTRACT

In an era marked by advancing technology, addressing communication barriers faced by individuals with speech and listening challenges is paramount. This study presents an innovative approach to facilitate seamless communication for those who rely on sign language. However, the problem is that not everyone learns sign language and learning takes time and effort and is sometimes discouraging also, Thinking of places where sign language could benefit this community when put in use, there are a few places that come to mind like Restaurants, stores, supermarkets, etc. If the business owner understands the problem and would adopt technology, which translates gestures into words/short sentences, this accessible experience would draw more customers and can gain more profit. Leveraging the capabilities of the volov5 architecture, this project endeavors to create an AI system capable of real time translation of sign language gestures into textual representation. In the proposed model the Images of sign language are captured using a webcam annotate, labelled and create YOLO format datasets for sign language The model implementation has the potential to break down communication barrier and facilitate smoother interaction in diverse setting. The system has undergone real-time testing and achieved. Best accuracy with reduced computational cost.

Keywords: **Sign language, Yolov5, Real time**

I. INTRODUCTION

Sign language serves as a valuable means of communication accessible to both the hearing and deaf communities The ability to communicate effectively is a fundamental aspects of the human experience. a restaurant named 'ISHAARA' in Mumbai, India, where owners are working for this deaf and mute community as they have hired hearing and speech impaired staff. The conversation happens like they have made customized menu cards wherein they have assigned particular signs for each cuisine. So, in this way, they are helping this community by providing jobs and making them feel equal in society. But this consumes more time and the process of communication may be slow. To make it more effective and time-saving, we have come up with this effective idea that would make work easier and can save time. The intersection of deep learning and computer vision has paved the way for transformative solutions, Among the deep learning frameworks, YOLO (You Only Look Once) has garnered substantial attention due to its exceptional capabilities in real-time object detection tasks. The YOLO approaches essence lies in its efficiency, as it performs object detection in a single forward pass through the network. YOLOV5 emerging as a promising framework for the development of accurate and efficient sign language recognition models. Sign language recognition is commonly associated with image understanding, consisting of two key phases: sign detection and sign recognition. Sign detection involves extracting specific features of an object based on particular parameters, while sign recognition entails identifying distinctive shapes that distinguish the object from others. To achieve this, the system will employ a webcam for image capture, and the preprocessing of signs will be executed within the Microsoft Visual Studio integrated development environment (IDE) using the OpenCV library. Notably, the prevailing trend in this field involves a growing reliance on machine learning, where machines are trained using diverse datasets to enhance their sign language recognition capabilities. In this project, the focus is on developing a Sign Language recognition system that can process input from image data using YOLOV5, a powerful Convolutional Neural Network (CNN). The system utilizes the neural network to extract relevant features and identify regions of interest, allowing it to accurately detect signs and translate them into text.

II. METHODOLOGY

The data flow diagram provide a clear overview of data movement, helping in understanding system functionality and communication. They are a vital tool for systems analysis, design, and documentation.

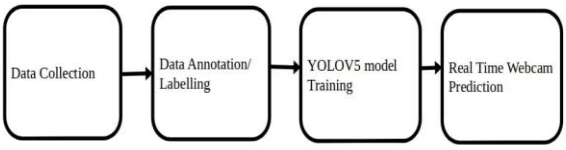

Fig 1 : Data Flow Diagram.

a. DATA COLLECTION:

Efficient training data collection is facilitated through tools such as Visual Studio Code (VSCode) and the Python programming language, leveraging the capabilities of a webcam. In this process, a custom dataset is curated by capturing sign images, including gestures like "Good Bye," "Yes," "No," "Reduce Price," and "Nice." A total of 1800 images, with 300 images captured for each sign, contribute to this dataset.

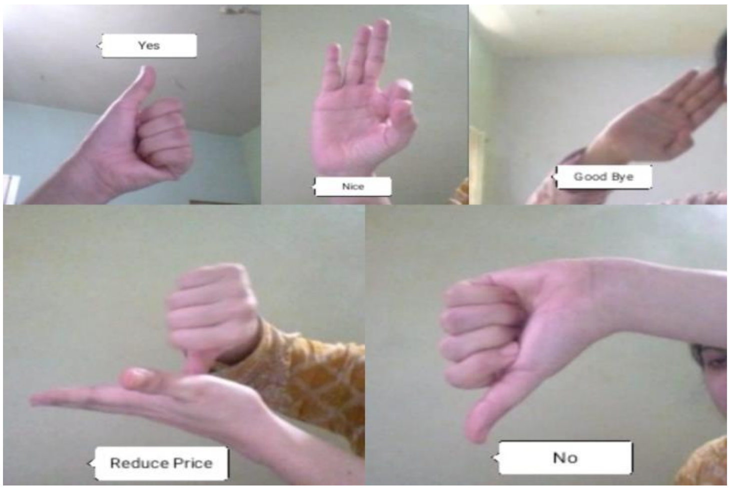

Fig 2 : Custom data set.

The dataset creation process involves a balanced approach, with 80% of the images allocated for training purposes and the remaining 20% for testing. This division ensures comprehensive model training and robust testing for accurate recognition. By integrating technology tools and meticulously capturing sign images, the training dataset's integrity and diversity are upheld, which is essential for achieving optimal object detection outcomes.

b. DATA ANNOTATION AND LABELLING:

Data annotation and labeling are essential components of preparing data for machine learning tasks, impacting the effectiveness of models across diverse applications like image recognition and natural language processing. The YOLO (You Only Look Once) network, a significant architecture in object detection, requires pre-labeled images for training, a process termed image annotation. This involves manually assigning object classes and outlining bounding boxes around objects within images. Yet, this task goes beyond manual input; it demands careful consideration of object classes and precise bounding box placement. The "make sense.ai" tool streamlines this intricate process by combining manual expertise with technological efficiency, yielding

PAGES: 188-192
9/10/23
JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY(JSRT) VOLUME-1 ISSUE-6 SEPTEMBER
Registered under MSME Government of India ISSN: 2583-8660

annotations that align with model requirements. YOLO's recognition capabilities hinge on annotations in a specialized format, encompassing object class and bounding box coordinates stored in text files. Each bounding box is defined by five parameters, encompassing object class, center coordinates, and box dimensions. This orchestrated interplay forms YOLO's distinctive annotations, driving effective object detection. In essence, data annotation merges human understanding with technological innovation, demonstrated by annotations that empower YOLO's recognition potential—an embodiment of meticulous craftsmanship that navigates the complexities of object detection.

c. TRAINING:

To initiate the model training process on Google Colab, ensure GPU (Graphic Processing Unit) utilization for enhanced performance. Start by cloning the YOLOv5 repository from the official GitHub source. Afterward, acquire all the necessary dependencies. The YOLOv5 model architecture requires meticulous configuration. This encompasses creating a tailored configuration file containing crucial specifications, including the number of classes (sign gestures) and opting for YOLOv5s architecture. Within this configuration file, establish links to the dataset's location and the annotation structure to facilitate model setup. The training phase involves executing the model over 300 epochs, with images sized at 640 and a batch size set at 8 for efficient training. Subsequent to training, The model's efficacy is assessed through rigorous evaluation conducted on a dedicated validation dataset. gauging its capability to accurately detect and classify sign language gestures.

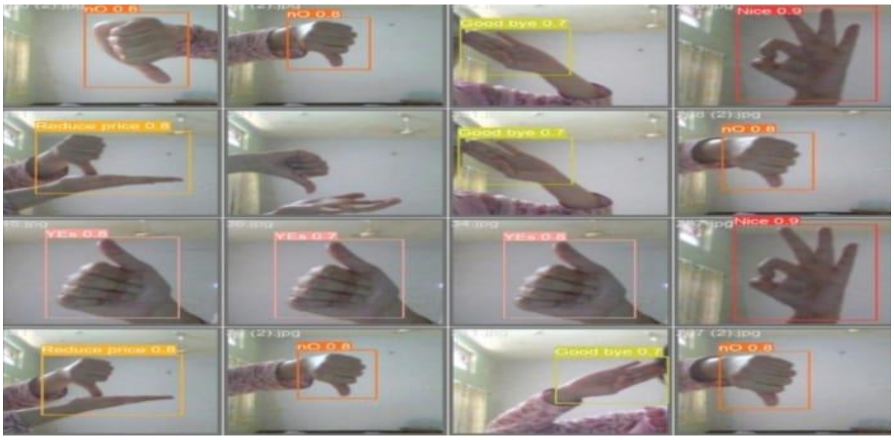

Fig 3 : Testing on Validation set.

III. RESULTS

After achieving convergence in training and demonstrating strong performance on the validation set, the model's suitability for real-world scenarios emerges. This pivotal phase involves subjecting the model to evaluation with authentic data, including live webcam input. Webcam frames are fed to the model for real-time predictions, enabling it to detect and classify sign language gestures. This process empowers the system to interpret and respond effectively to user interactions, signifying a transformative synergy between technology and human expression that holds the promise to bridge communication gaps and enhance accessibility.

PAGES: 188-192
9/10/23
JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY(JSRT) VOLUME-1 ISSUE-6 SEPTEMBER
Registered under MSME Government of India ISSN: 2583-8660

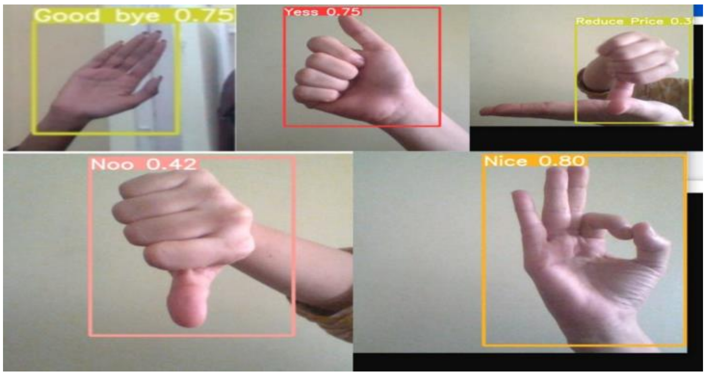

Fig 4 : Real Time Webcam Prediction.

During the training process, maintaining vigilant oversight over loss values and performance metrics is imperative. This continuous evaluation offers valuable insights into the model's advancement and its capacity for learning and adaptation.

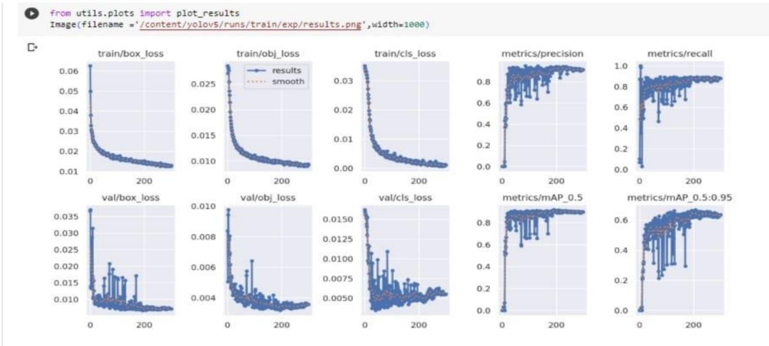

Fig 5 : Graphs of Validation, Accuracy, Loss and Training.

Training a custom dataset using YOLOv5 demands a thorough evaluation of data quality, model architecture, and hyperparameters. This comprehensive approach is essential to attain highly accurate results in object detection.

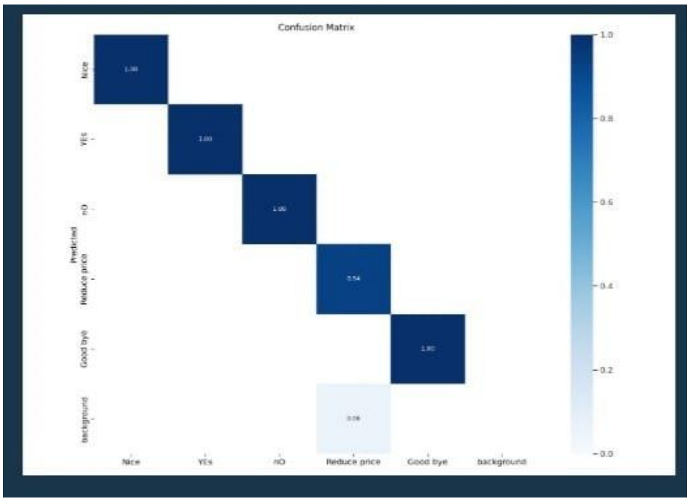

Fig 6 : Confusion Matrix

IV. CONCLUSION

The sign language project utilizing YOLOV5 demonstrates a promising application of computer vision technology, facilitating real-time detection and recognition of various sign language gestures. By facilitating effective communication between individuals who are deaf and those who are not proficient in sign language, we bridge a significant linguistic and cultural divide, the project offers valuable support to the deaf and hard of hearing community. The model's speed and accuracy in identifying sign language gestures from live video streams or recordings contribute to a seamless user experience. However, ongoing challenges include the need for a diverse and expansive training dataset, continuous data collection, and user feedback for fine-tuning and optimization. By incorporating user- friendly interfaces and accessibility features, the project aims to ensure inclusivity and enhance the recipient's understanding of translated messages. Overall, this endeavor represents a significant step towards fostering and accessibility, potentially breaking down communication barriers and empowering individuals with diverse communication needs, making a meaningful impact on society's connectivity and compassion.

REFERENCES

1. X. Jiang and W. Ahmad, "Hand Gesture Detection Based Real-Time American Sign Language Letters Recognition using Support Vector Machine," 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), Fukuoka, Japan, 2019, pp. 380-385, doi: 10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00078.
2. K. Nimisha and A. Jacob, "A Brief Review of the Recent Trends in Sign Language Recognition," 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2020, pp. 186-190, doi: 10.1109/ICCSP48568.2020.9182351.
3. A. B. Jani, N. A. Kotak and A. K. Roy, "Sensor Based Hand Gesture Recognition System for English Alphabets Used in Sign Language of Deaf-Mute People," 2018 IEEE SENSORS, New Delhi, India, 2018, pp. 1-4, doi: 10.1109/ICSENS.2018.8589574.
4. M. J, B. V. Krishna, S. N. S and S. K, "Sign Language Recognition using Machine Learning," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICSES55317.2022.9914155.
5. J. J. Raval and R. Gajjar, "Real-time Sign Language Recognition using Computer Vision," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp. 542-546, doi: 10.1109/ICSPC51351.2021.9451709.
6. A. Er-Rady, R. Faizi, R. O. H. Thami and H. Housni, "Automatic sign language recognition: A survey," 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Fez, Morocco, 2017, pp. 1-7, doi: 10.1109/ATSIP.2017.8075561.
7. Davi Hirafuji Neiva, Cleber Zanchenttin, Gesture recognition: A review focusing on sign language in a mobile context, Expert Systems with Applications, 103, 159-183 (2018).
8. Suharjito, Ricky Anderson, Fanny Wiryana, Meita Chandra Ariesta, Gede Putra Kusuma, Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output, 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, 3-14 October 2017, Bail, Indonesia (2017).
9. Ming Jin Cheok, Zaid Omar, Mohamed Hisham Jaward, A review of hand gesture and sign language recognition techniques (2017).
10. Tejashri J. Joshi, Shiva Kumar, N. Z. Tarapore, Vivek Mohile, Static Hand Gesture Recognition using an Android Device, in International Journal of Computer Applications (0975 – 8887) , Vol 12, No.21 (2015).
11. 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.