Yogesh Pawar, 2025, 13:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

SignMate: Sign Language Detection System

Yogesh Pawar, Mrudula Kotgire, Vihar Motghare, Riyah More, Om More Mrinaalini Shankar
Department of Engineering, Sciences and Humanities (DESH)
Vishwakarma Institute of Technology, Pune, Maharashtra, India

Abstract- This paper outlines the designing and improvement of a real-time sign language recognition system that can precisely translate hand movements from a typical webcam. The basis of the system is YOLOv5 (Ultralytics), a cutting-edge object detection model, utilizing the PyTorch deep learning framework for implementation and training. Data acquisition entailed capturing personal Indian Sign Language (ISL) gestures with a webcam under normal illumination. These movements were carefully annotated and exported with Roboflow to produce a high-quality, YOLO-ready dataset. The system initially supported 7 sign classes. Through dramatic enhancements, the dataset grew to 15 classes: "Bye," "Congratulations," "How_Are_You," and "No_Worries," with 50-100 images taken per class to maximize diversity and balance. The YOLOv5s model was trained for 30 epochs with 16 batch size and an image input of 416x416 and achieved a remarkable average mAP@0.5 of more than 98%. Real-time inference is conducted using the detect.py script with accurate bounding box predictions and confidence scores. The whole project was implemented in Python, using Visual Studio Code and Anaconda for managing the environment, and is run locally on a CPU with OpenCV for video processing. The system thus shows much better real-world applicability, diversity of classes, and usability than its previous version and represents a great leap toward more effective gesture-based communication. Future projects involve building an Android app through a Flask API, enhancing the graphical user interface, testing out more larger YOLOv5 models, and incorporating temporal tracking for video-based identification.

Keywords-Sign Language Detection, YOLOv5, Real-time Object Detection, PyTorch, Hand Gesture Recognition, Machine Learning, Computer Vision, Roboflow

I. INTRODUCTION

FOR the deaf and hard-of-hearing communities, sign language is an essential communication tool, so reliable and user-friendly interpretive technologies are required. This project creates an advanced hand gesture detection system to meet the urgent need for real-time sign language interpretation. This system's core is a specially trained object detection model that can recognize several hand gesture signs from a standard webcam feed, greatly improving accessibility and gesture-based communication.

Our system achieves high accuracy and real-time performance by utilizing cutting-edge machine learning platforms and tools. In particular, the project's core machine learning platform is PyTorch, an open-source deep learning framework. We used YOLOv5 (Ultralytics), a well-known real-time object detection model renowned for striking a balance between speed and accuracy, as the detection engine. To create a well-labeled, YOLO-compatible dataset, custom Indian Sign Language (ISL) gestures were painstakingly captured using a webcam in natural lighting, then annotated and exported using Roboflow. The YOLOv5 model was then trained using this anotated dataset. Using Visual Studio Code and Anaconda for thorough package and virtual environment management, the entire development environment was set up in Python. With PyTorch enabling inference and OpenCV handling video handling, the model operates locally on a CPU. A seven-class model was the main focus of the initial development, but it was later improved and broadened to correctly identify fifteen different sign language classes, indicating a notable increase in both scope and usefulness.

II. LITERATURE REVIEW

The field of sign language recognition has seen significant advancements through computer vision and deep learning techniques. Key contributions from existing research include:

- Review of Gesture Recognition (Qi et al., 2024): Reviewed vision-based hand gesture recognition methods for human-robot interaction (HRI). This work compared sensor and camera approaches and highlighted the need for more real-time analysis and optimization for dynamic environments.
- Deep Learning Approach for ISL (Autee & Raut, 2020): Utilized a CNN-based model for real-time Indian Sign Language (ISL) gesture recognition. This approach was effective for static signs but showed limitations in handling dynamic gestures.
- MediaPipe and CNNs for ASL (Kumar et al., 2023): Presented a real-time ASL recognition system using MediaPipe and CNNs for alphabet classification. It achieved 99.95% accuracy on a large-scale dataset of 26 ASL letters and was noted for creating lightweight, hardware-independent solutions.
- Deepsign: LSTM+GRU for ISL (Kothadiya et al., 2022): Introduced "Deepsign," a deep learning model combining LSTM and GRU for ISL gesture recognition. It achieved 97% accuracy using the ISL20 dataset and improved temporal pattern recognition for real-world ISL inputs.
- Real-time Indian Sign Language Recognition (Shenoy et al., 2018): Developed a real-time ISL recognition system using grid-based features. It achieved high accuracies (99.7% for poses, 97.23% for gestures) and was implemented as an Android app with a server.
- Continuous Sign Language Recognition with MediaPipe Holistic (Srivastava et al., 2024): Proposed a real-time ISL recognition system using MediaPipe Holistic and LSTM, achieving 88.23% accuracy for 45 gestures. This work suggested the need for more data and expansion to broader sign languages and real-world conditions.
- Fine-tuning Sign Language Recognition Models (Novopoltsev et al., 2023): Explored cross-lingual fine-tuning for SLR using VideoSWIN and MViT models, achieving real-time CPU performance. While demonstrating effective transfer learning, it highlighted a lack of continuous SLR evaluation and potential for broader language support and mobile deployment.

III. METHODOLOGY/EXPERIMENTAL

Synthesis/Algorithm/Design/Method

In our project SignMate we use YOLOv5 which is You Only Look Once version 5 small. It ia a object detection model which is helpful for fast and accurate real time gesture recognition.

Mode Architecture:

Backbone: CSPDarknet – which brings out essential visual features.

Neck: PANet- improvises fusion feature

Head: YOLO Layer: for bounding boxes and a class probability.

Dataset Creation:

15 different signs were captured one by one using web camera.

50-100 images per gesture were captured under different conditions like background or light.

Roboflow was used to make bounding box for each sign.

Preprocessing:

Image resolution was changed to 416 x 416

Dataset was balanced to make it more generalized.

Training:
Training was done in pytorch with use of YOLOv5s training pipeline
Parameters: image size, epochs(30), batch size(16).
Dataset was split in 80% and 20%. 80% for training and 20% for validation.
Inference
Detection was done using detect.py script.
Signs were identified by bounding boxes around it and precision score on web camera interface.

IV. RESULTS AND DISCUSSIONS

The Indian Sign Language (ISL) detection system was rigorously tested using a custom-built dataset containing 15 unique gesture classes. Each image in the dataset was accurately labeled using Roboflow and utilized to train a YOLOv5 model implemented in PyTorch. The model's performance was assessed using standard object detection metrics: mean Average Precision (mAP), Precision, and Recall.
A. Initial Evaluation of the Model
In the first evaluation phase, the model demonstrated the following results:
mAP@0.5: 0.983
Average Precision: 0.96
Average Recall: 0.94
Live testing through webcam input highlighted the model's strong capability to detect commonly used gestures such as "Hello", "Yes", "No", and "Thank You". However, some gestures, including "Okay", "Peace", and "Great", occasionally caused confusion due to their visual similarity to "Hello". These early misclassifications were largely the result of limited image diversity within those gesture categories.
Enhancing Accuracy through Dataset Augmentation
To minimize such misclassifications, a focused dataset augmentation strategy was employed. An additional 50–100 images were collected for each of the visually similar or problematic gestures. This enriched dataset was then used to fine-tune the YOLOv5 model by leveraging its pre-trained weights. The fine-tuning process resulted in a noticeable improvement in detection performance, particularly for gestures that were previously ambiguous.
Additionally, to maintain ethical standards in AI development, any gesture that had the potential to be offensive or misinterpreted was carefully reviewed and permanently excluded from the dataset. This proactive measure supports the development of a responsible and inclusive AI system suitable for real-world use.
Real-time Detection Performance and Future Enhancements
The finalized model delivered reliable real-time performance when run using the detect.py script, handling live webcam input with low latency and maintaining high accuracy across most gesture classes. This stability and responsiveness underscore the model's potential for use in real-time accessibility tools.
Moving forward, future enhancements will aim at integrating the model into graphical user interfaces (GUIs) and mobile applications. Key focus areas include optimizing the system for low-resource environments and creating intuitive user experiences, thereby broadening its accessibility and impact.

V. OTHER RECOMMENDATIONS

To further improve the sign language recognition system, future development can include the addition of background removal techniques to reduce recognition errors in noisy backgrounds. Reducing the time delay between input gestures and output text is also important, as the current system requires each gesture to be held for a few seconds.

Expanding the dataset with more diverse images—covering different lighting conditions, backgrounds, and hand variations—can enhance model accuracy and generalization. Adding a graphical user interface or developing a mobile application would also make the system more user-friendly and accessible.

Additionally, other or combined models other than CNNs may further enhance the performance, particularly for continuous or real-time gesture recognition.

VI. FUTURE SCOPE

There is potential for further development of the Real Time Indian Sign Language Recognition System with YOLOv5.

Further Constructive Enhancements.
Improvement of the dataset through New Class Incorporation: Future works should focus on signing systems extending the current 13-class dataset to include regional signs at the interchange of simple signs, compound signs, and complex sentences. Capturing data from different individuals, lighting backgrounds, and various camera angles of themselves would improve the model's robustness and generalization.

Improvement of Model Architecture: Apart from its great performance, with mAP@0.5 of 0.983 doing they YOv5 was impressive, exploring new architectures such as YOLOv8, YOLO-NAS, and transformer based models could deliver higher precision and speed of inference. High efficiency can be expected with the implementation of attention mechanism multi-scale feature extraction for detecting subtle differences in the gestures.

Enhancing frame-per-second (FPS) metrics while maintaining the same features voided during real-time system execution is yielding to model quantization, pruning, and knowledge distillation. These techniques allow access to further improvement of performance without functionality impact which is important for resource demand devices such as smartphones and embedded systems.

Translating Signs to Speech Systems: The collaboration with text-speech synthesis units can complete the communication interface by producing sound for recognized gestures in sign language. This two-way communication system would improve accessibility to a great extent.

Teaching Tools: The interactive education development for teaching sign language in the form of games with tracking features issued rewards will help hearing persons as well as the deaf learn the Indian Sign Language.

Integration and Expandability
Embedding the detection system into smart IoT devices, public kiosks, and digital signage can make the environments more accommodating to all people. The inclusion of virtual assistants and smart appliances can allow for gesture-based control systems.

Deploying AR/VR technologies for sign language teaching can facilitate active learning for users as well as for remote communication between sign language users and non-sign language users.
Integrating gesture recognition with expression and lip reading brings a new level to communication systems by deepening the understanding of sign language and creating new systems capable of capturing primary channels of sign communication.
Research Directions
Systems capable of adapting to new gesture learning continuously are called continuous learning systems.

VII. CONCLUSION

This research demonstrates the development of an Indian Sign Language detection system using YOLOv5 achieving a mean average precision of 0.983 across 13 gesture classes. This project addresses an unfulfilled gap in the existing literature towards the need for real-time automated sign language recognition systems in India; thus, advancing further towards the digital divide and communication technologies.

The step-by-step methodology followed in this work starting from creating custom datasets to refining the models applies the concepts of computer vision engineering to sign language recognition. Real-time gesture recognition accuracy and computational efficiency for deployment aimed application was best achieved using the YOLOv5 architecture.
This work has made significant contributions through developing a webcam integrated real-time system which recognizes and tracks a wide range of Indian Sign Language gestures, compiled detailed datasets for frequent use gestures, and demonstrating multi-class, high-precision recognition performance. The flexibility of the approach to maintain performance while

incorporating additional gestures into the model indicates its adaptability to iterative enhancement strategies.

ACKNOWLEDGMENT

Our profound appreciation goes out to Vishwakarma Institute of Technology, Pune, for providing the tools and assistance required for this study. A special thank you to Prof. Yogesh Pawar, our guide, for all of his helpful advice and perceptive criticism during the process. Furthermore, we thank to our group members for your efforts to system development and testing.

.

REFERENCES

1. D. Kaur and M. Kaur, “Sign Language Recognition System for Deaf and Dumb People,” International Journal of Engineering Trends and Technology, vol. 4, no. 3, pp. 102–106, 2013.
2. Juneja S, Juneja A, Dhiman G, Jain S, Dhankhar A, Kautish S. computer Vision-Enabled character recognition of hand Gestures for patients with hearing and speaking disability. Mobile Information Systems. 2021
3. Yulius Obi, Kent Samuel Claudio, Vetri Marvel Budiman, Said Achmad, Aditya Kurniawan, Sign language recognition system for communicating to people with disabilities, Procedia Computer Science, Volume 216, 2023,
4. O. Yilmaz and H. Y. Keles, "Real-time sign language recognition based on YOLO algorithm," Neural Comput. Appl., vol. 36, no. 8, pp. 4269-4285, Mar. 2024
5. Kadhim RA, Khamees M. A real-time American Sign Language recognition system using convolutional neural network for real datasets. Tem Journal. 2020.
6. A. Mehta, R. Shivagonde, and A. Joglekar, "Indian Sign Language Recognition System," ResearchGate, 2017. [Online]. Available:
7. S. Sharma and K. Kumar, "Indian Sign Language recognition system using SURF with SVM and CNN," Array, vol. 14, p. 100121, Jul. 2022. [Online]. Available:
8. P. Singh, A. Kumar, and R. Sharma, "Indian Sign Language Character Recognition System," in Proc. IEEE Int. Conf. Electron. Commun. Inf. Technol., Kharagpur, India, 2023, pp. 1-6. [Online]. Available:
9. A. Khan, M. N. Cheema, and A. Akram, "Deepsign: Sign Language Detection and Recognition Using Deep Learning," Electronics, vol. 11, no. 11, p. 1780, May 2022. [Online]. Available:
10. M. A. Rahman, M. K. Hasan, and S. Islam, "American Sign Language Detection using YOLOv5 and YOLOv8," Research Square, preprint, Jul. 2023. [Online]. Available:
11. O. Yilmaz and H. Y. Keles, "Real-time sign language recognition based on YOLO algorithm," Neural Comput. Appl., vol. 36, no. 8, pp. 4269-4285, Mar. 2024. [Online]. Available:
12. A. Wijaya and D. Sari, "Sign Language Detection System Using YOLOv5 Algorithm," Sci. J. Inform vol. 11, no. 1, pp. 89-98, Feb. 2024. [Online]. Available:
13. T. Ahmed and S. Khan, "Real Time American Sign Language Detection Using Yolo-v9," arXiv preprint arXiv:2407.17950, Jul. 2024. [Online]. Available