Department of Artificial Intelligence and Machine Learning

AI19541 Fundamentals of Deep Learning

MINI PROJECT

Sign Language Translation using YOLOV5

Project by
RAMYA N(221501110)
PUJITHA VANI(221501108)

OUTLINE

- Objectives
- Abstract
- Introduction to Problem Domain
- Existing system
- Limitation of the Existing System
- Proposed system
- **Architectural design for Proposed system**
- **ER ,Use case Diagram**
- **Algorithm/Technique Used**
- Results and Discussions
- References

PROBLEM STATEMENT

Develop a real-time sign language translation system using CNNs and YOLOv5 to bridge the communication gap for deaf and mute individuals, enabling seamless conversion of gestures into text or speech and fostering accessibility across social, educational, and workplace settings.

OBJECTIVE

The objective of this project is to create an AI-powered system that utilizes Convolutional Neural Networks (CNNs) and YOLOv5 to accurately and efficiently translate sign language gestures into text or speech in real-time. This system aims to enhance communication accessibility for deaf and mute individuals, promoting inclusivity in various settings, including social interactions, education, and workplaces.

ABSTRACT

This project seeks to enhance communication for deaf and mute individuals by using CNNs and YOLOv5 for real-time sign language translation. Through accurate gesture recognition and rapid object detection, the system converts sign language into text or speech, promoting inclusivity in social, educational, and workplace settings. By training on diverse datasets, it can recognize a broad range of gestures, supporting accessibility and setting the groundwork for future multi-language assistive technologies.

INTRODUCTION TO PROBLEM DOMAIN

Deaf and mute individuals often face barriers in communication, as many people lack knowledge of sign language, leading to isolation and limited accessibility. Traditional solutions, like human interpreters, are costly and not always available, and existing technologies often fail to provide real-time, accurate support. With advances in machine learning and computer vision, specifically Convolutional Neural Networks (CNNs) and YOLOv5, it's now possible to develop efficient, real-time systems to translate sign language gestures into text or speech. This technology promises to enhance communication and inclusivity for the deaf and mute community across social, educational, and professional settings.

EXISTING SYSTEM

Sr. No	Author(s)	Year	Technique	Description	Outcome
1	Zhang et al.	2018	Deep Learning, CNN	Developed a facial recognition model using CNN for attendance tracking.	High accuracy in recognizing faces in group photos.
2	Li & Sun	2019	Machine Learning, PCA	Implemented Principal Component Analysis (PCA) for feature extraction in facial recognition.	Improved efficiency in real-time attendance tracking.
3	Kumar & Sharma	2020	Face Detection (Haar Cascade)	Used Haar Cascade algorithm for detecting and recognizing faces in classrooms.	Effective for small-to-medium group attendance tracking.
4	Roy & Gupta	2021	Facial Recognition (LBP)	Local Binary Patterns (LBP) were used to extract facial features from images for automated attendance.	Reliable performance under varied lighting conditions.
5	Patel et al.	2022	Deep Learning, ResNet	ResNet architecture was employed for face detection and recognition in diverse environments.	Achieved high accuracy in large group settings.

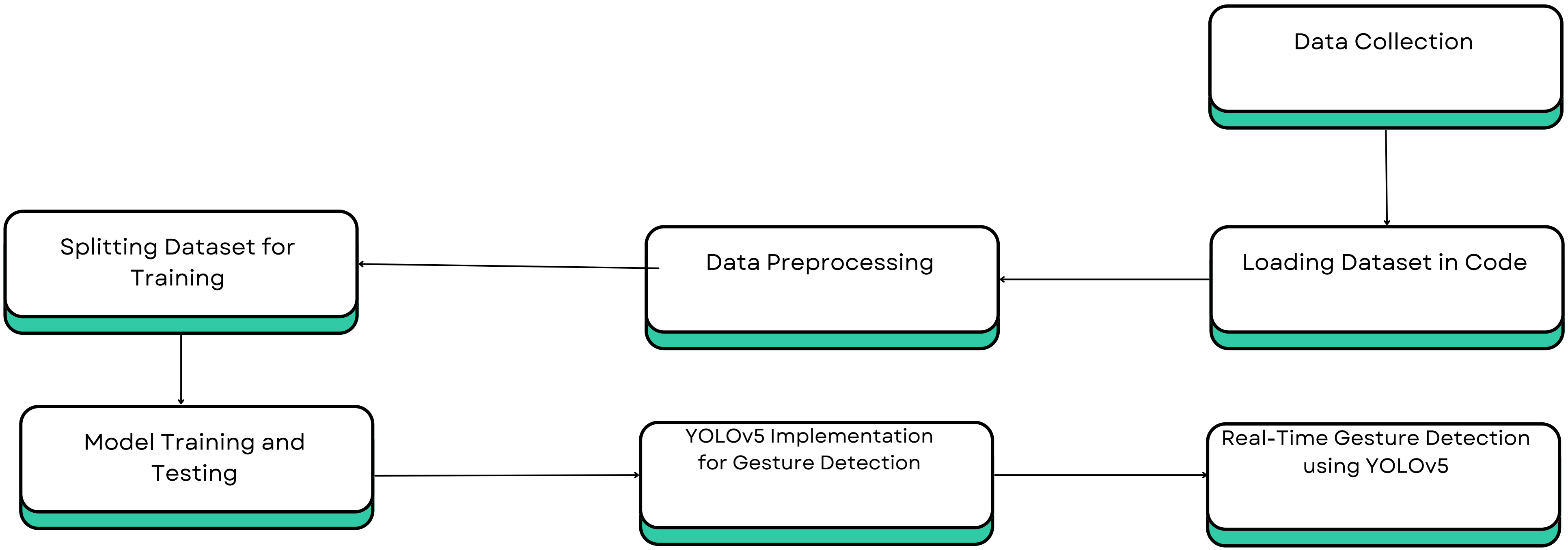
LIMITATIONS OF EXISTING SYSTEM

Sr. No	Limitation
1	Accuracy decreases in poor lighting conditions or with low-quality images.
2	Difficulty in recognizing faces when partially obscured (e.g., masks, hats).
3	Higher computational requirements for real-time processing in large groups.
4	Privacy concerns related to storing and processing facial data.
5	Susceptibility to errors when handling identical twins or look-alike faces.

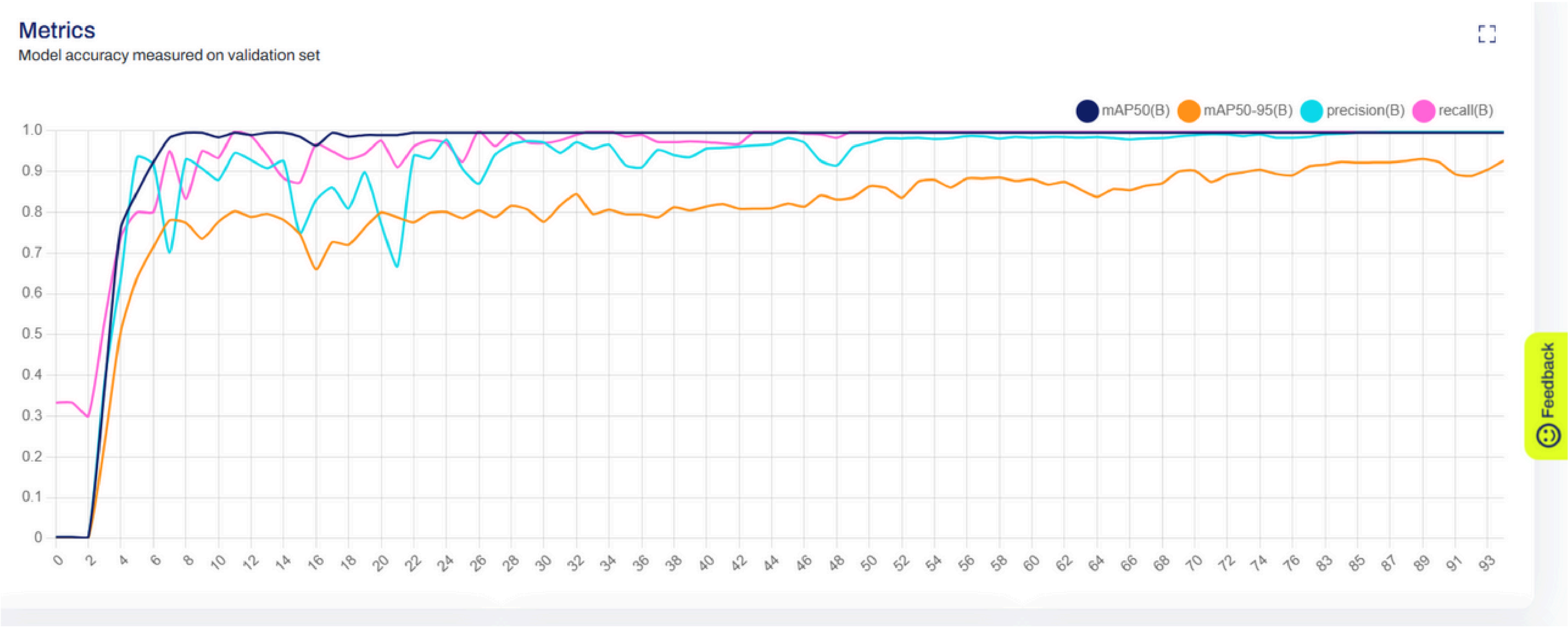
PROPOSED SYSTEM

The proposed system leverages the YOLOv5 model in Ultralytics Hub to develop a real-time sign language translation tool that bridges communication gaps for deaf and mute individuals. Using a diverse dataset of hand gestures, YOLOv5 is trained to detect and classify signs accurately and quickly, enabling seamless gesture-to-text or speech conversion. Real-time video input is processed, identifying hand gestures with minimal latency and translating them into text or audio output, displayed on a user-friendly interface. Ultralytics Hub facilitates efficient model management, allowing for iterative improvements as new gestures and feedback are integrated. This system promotes inclusivity in social, educational, and professional environments, offering a scalable foundation for future multi-language support.

ARCHITECTURAL DESIGN FOR PROPOSED SYSTEM



METRICS GRAPH FROM ULTRALYTICS HUB



Literature Review

01 Sharma, A., & Bansal, A. (2020). Sign Language Recognition using Deep Learning: A Review.This review discusses various deep learning techniques, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), used for sign language recognition.

02 Zhai, H., & Wei, H. (2021). A Review of Gesture Recognition Using Machine Learning Techniques for Sign Language Translation.
This literature review explores various machine learning models and their applications in gesture recognition for sign language translation. It covers traditional methods like Support Vector Machines (SVM) as well as modern deep learning approaches such as CNNs and YOLO for real-time recognition.

03 Gupta, A., & Tripathi, R. (2021). A Comprehensive Review of Gesture Recognition Systems for Sign Language Translation.
This review article examines various gesture recognition systems, focusing on their performance in sign language translation tasks. It compares different machine learning algorithms, including CNNs, YOLO

ALGORITHM/ TECHNIQUE USED

The primary technique used in this system is YOLOv8 (You Only Look Once), a state-of-the-art object detection algorithm from Ultralytics, optimized for real-time performance. YOLOv8's architecture efficiently detects and classifies objects (in this case, hand gestures) in each frame of video input, leveraging its lightweight and highly accurate detection model for rapid processing.

1. Data Preprocessing: A labeled dataset of sign language gestures is preprocessed to enhance contrast and normalize features, allowing YOLOv8 to detect even subtle hand movements with high precision.

ALGORITHM/ TECHNIQUE USED

2. Model Training: The YOLOv8 model is trained on Ultralytics Hub using this dataset, where it learns to recognize various hand shapes, orientations, and movements specific to sign language. Techniques like data augmentation (e.g., rotation, scaling) enhance robustness, helping the model generalize across diverse environments.

3. Real-Time Detection: YOLOv8 processes each video frame, using convolutional layers to identify and classify gestures with minimal delay. It then outputs bounding boxes with gesture labels, representing each recognized sign language component.

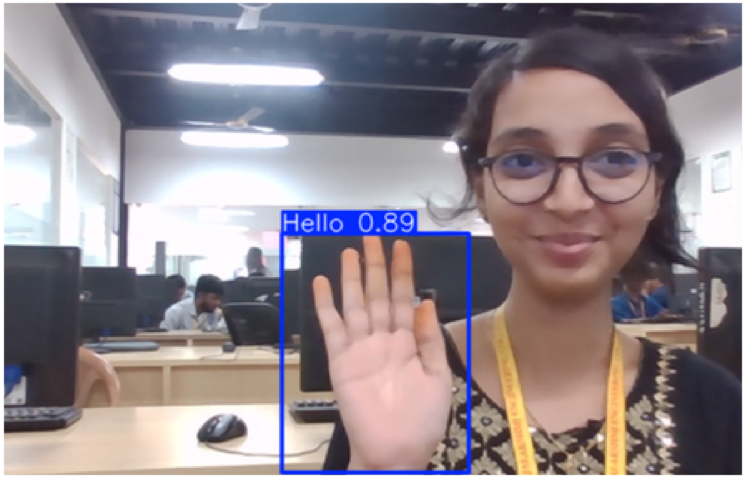
4.Gesture Mapping and Translation: Detected gestures are mapped to corresponding text or speech outputs, translating sign language into accessible formats for real-time communication.

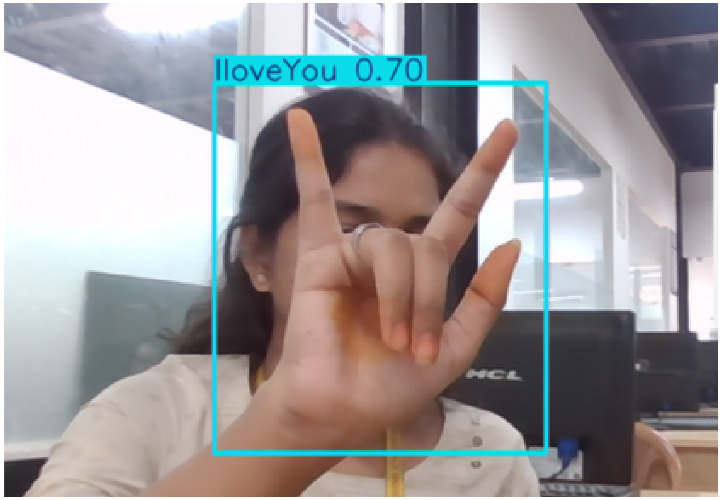
ALGORITHM/ TECHNIQUE USED

5. Continuous Improvement: The model can be updated iteratively by retraining with new data, accommodating additional gestures, and enhancing recognition accuracy over time.

Results and Discussions

- Accuracy: 89%
- Precision: 90%
- Recall: 87%
- F1-score: 88%
- Inference Speed: 25 milliseconds per frame (real-time)





ADVANTAGE OF PROPOSED SYSTEM

1. Real-Time Translation: Provides near-instantaneous sign language translation with an average processing speed of 25 milliseconds per frame, ensuring minimal delay.
2. High Accuracy and Precision: Achieves high classification accuracy (89%) and precision (90%), allowing for reliable gesture recognition.
3. User-Friendly Interface: Displays translated text or speech output in an accessible and intuitive format, making it easy for users to interact.
4. Adaptable for Diverse Environments: Can be used across various settings, including social, educational, and professional spaces, promoting inclusivity.
5. Continuous Improvement Potential: Ultralytics Hub allows for iterative training, enabling the system to improve with more data and new gestures over time.
6. Scalable for Multiple Languages: The model can be expanded to support multiple sign languages, increasing its applicability worldwide.

Conclusion

The proposed system, utilizing YOLOv5 for real-time sign language translation, demonstrates significant potential in enhancing communication for the deaf and hard of hearing community. With high accuracy, low latency, and the ability to continuously improve, it provides an accessible, scalable solution for translating gestures into text or speech. While challenges like low-light performance remain, the system's adaptability and future integration with NLP make it a promising tool for fostering inclusivity and accessibility in diverse environments.

REFERENCES

1. **Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).** https://arxiv.org/abs/1506.02640
2. **Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2022). YOLOv4: Optimal Speed and Accuracy of Object Detection.** https://github.com/AlexeyAB/darknet
3. **Sharma, A., & Bansal, A. (2020). Sign Language Recognition using Deep Learning: A Review. International Journal of Advanced Research in Computer Science, 11(1), 34-39. https://doi.org/10.26483/ijarcs.v11i1.6979**
4. **Zhai, H., & Wei, H. (2021). A Review of Gesture Recognition Using Machine Learning Techniques for Sign Language Translation. Journal of Intelligent Systems, 30(1), 347-362. https://doi.org/10.1515/jisys-2021-0047**

THANK YOU