PHONEVISION: TEACHING MACHINES TO SEE AND READ PHONE NUMBERS

Mrs. Divya M,
Department of CSE
Rajalakshmi Engineering College Chennai, India
divya.m@rajalakshmi.edu.in

Sandhiya K,
Department of CSE
Rajalakshmi Engineering College Chennai, India
220701244@rajalakshmi.edu.in

ABSTRACT - The growing need for accessible communication tools for the deaf and mute community has led to significant advancements in sign language translation technology. This project introduces a real-time Sign Language Translation System utilizing the YOLOv5 model, renowned for its efficient hand gesture detection capabilities. By integrating YOLOv5 with Convolutional Neural Networks (CNNs), the system achieves accurate gesture recognition and provides seamless translations into text or speech, making everyday interactions more inclusive and accessible.

Designed to handle diverse sign languages, the system supports real-world scenarios by ensuring fast and precise translations, thereby fostering inclusive communication between sign language users and non-signers. By combining high accuracy with real-time performance, the proposed system highlights the potential of leveraging deep learning for accessible and adaptive communication.

In parallel, facial recognition technology has seen widespread application, particularly in security and automated attendance systems. Notable methods, such as Haar Cascade Classifiers and OpenCV-based implementations, have proven effective for real-time face detection under varying conditions. Future enhancements for the proposed gesture recognition system include expanding support to multiple sign languages and optimizing model performance for broader adaptability across different environments.

Sign language is a vital mode of communication for individuals with hearing or speech impairments. Despite its importance, there remains a significant communication barrier between sign language users and the majority of people who are not familiar with it. Bridging this gap requires an intelligent and efficient system capable of translating sign language into spoken or written language in real-time. This project addresses that challenge through the development of a Sign Language Translator using YOLOv5, a powerful object detection model that combines speed and accuracy, making it ideal for real-time applications.

The proposed system uses a camera as an input device to capture real-time video frames. YOLOv5, a one-stage object detection framework, is employed to identify and classify static or dynamic hand gestures. The model is trained on a custom dataset consisting of sign language alphabets, digits, or words, with annotated bounding boxes indicating the region of interest (the hand gestures). After detection, the model outputs the corresponding label for each gesture, which is then mapped to the appropriate character or word.

YOLOv5 is chosen for its lightweight architecture, high inference speed, and excellent performance on embedded systems and edge devices. Compared to traditional CNN-based classifiers or older object detection frameworks like R-CNN or SSD, YOLOv5 allows for real-time gesture recognition without compromising detection accuracy. Its modularity also enables customization and fine-tuning for specific sign language variants, such as ASL (American Sign Language) or ISL (Indian Sign Language).

The system is designed to work in diverse lighting and background conditions, enhancing its robustness in real-world environments. To achieve this, the dataset includes gesture samples taken under varying angles, lighting scenarios, and hand orientations. Data augmentation techniques such as rotation, scaling, and flipping are applied during training to increase the model's generalizability and reduce overfitting.

Post-processing steps are used to refine detection results and eliminate false positives. Once a gesture is recognized, it is translated into text, which can optionally be converted to speech using a Text-to-

Speech (TTS) engine. This multi-modal output ensures accessibility for both visual and auditory users.

To improve usability, a user-friendly interface is developed using Python (with libraries such as OpenCV, PyTorch, and Tkinter) or integrated into a mobile application using platforms like Android Studio. The interface displays live video, recognized signs, and the resulting translations in real-time. Future versions of the system can include support for continuous gesture recognition, dynamic signs (involving motion), and full sentence translations using Natural Language Processing (NLP).

Keywords - Facial recognition, Attendance management, Haar Cascade Classifier, OpenCV, Real-time detection, Face detection and training, Automated attendance system, Computer vision, Image processing, Machine learning, Customizable facial recognition, Attendance tracking system

I. INTRODUCTION

The growing demand for accessible communication tools for individuals with hearing and speech impairments has driven the development of intelligent sign language translation systems. Traditional methods of communication for the deaf and mute community often require human interpreters, which may not be readily available in all situations. This project addresses this gap by introducing a modular, real-time Sign Language Translation System that leverages the capabilities of YOLOv5 and Convolutional Neural Networks (CNNs) for efficient and accurate hand gesture recognition.

The system architecture consists of multiple integrated modules to ensure robust performance. It begins with a Data Collection module that captures a variety of sign gestures through video or image input. The Data Preprocessing stage prepares this data for training by resizing and labeling it appropriately. The Model Training phase utilizes YOLOv5 for rapid detection of hand gestures, and CNNs for gesture classification, resulting in a
trained model capable of recognizing gestures in real time.

Upon deployment, the Prediction Module operates in real-time to detect and translate gestures into text or speech outputs, thereby facilitating seamless communication between sign language users and non-signers. The Backend ensures efficient processing with low latency and stores recognized gestures for future improvements.

Furthermore, the system incorporates an automated attendance tracking feature using facial recognition, powered by Haar Cascade Classifiers and OpenCV. This additional functionality makes the system a comprehensive tool not only for inclusive communication but also for use in institutional and organizational environments requiring reliable and efficient attendance monitoring.

By combining deep learning with computer vision, the proposed system offers a scalable, cost-effective, and user-friendly solution that promotes inclusivity and accessibility. Its modular design allows easy expansion, such as supporting multiple sign languages or adapting to new environments, thus paving the way for broader societal impact. Communication is a fundamental human need, yet millions of people around the world who rely on sign language often face daily barriers when interacting with those who don't understand it. In a world increasingly powered by technology, there remains a gap in accessibility for the deaf and hard-of-hearing community. Bridging this gap requires more than empathy—it demands innovation.

This project introduces a smart solution that leverages the power of computer vision and deep learning to create a real-time Sign Language Translator using YOLOv5, one of the fastest and most efficient object detection models available today. By recognizing hand gestures through a camera and instantly converting them into readable text or spoken language, the system aims to break down communication walls and promote inclusivity.

Unlike traditional gesture recognition systems that rely on gloves or sensors, this solution is vision-based, contactless, and adaptable for everyday use on mobile and desktop platforms. It is designed not just as a technical project, but as a meaningful step toward making communication more accessible and

universal. Sign language is a vital mode of communication for individuals with hearing or speech impairments. Despite its importance, there remains a significant communication barrier between sign language users and the majority of people who are not familiar with it. Bridging this gap requires an intelligent and efficient system capable of translating sign language into spoken or written language in real-time. This project addresses that challenge through the development of a Sign Language Translator using YOLOv5, a powerful object detection model that combines speed and accuracy, making it ideal for real-time applications.

The proposed system uses a camera as an input device to capture real-time video frames. YOLOv5, a one-stage object detection framework, is employed to identify and classify static or dynamic hand gestures. The model is trained on a custom dataset consisting of sign language alphabets, digits, or words, with annotated bounding boxes indicating the region of interest (the hand gestures). After detection, the model outputs the corresponding label for each gesture, which is then mapped to the appropriate character or word.

YOLOv5 is chosen for its lightweight architecture, high inference speed, and excellent performance on embedded systems and edge devices. Compared to traditional CNN-based classifiers or older object detection frameworks like R-CNN or SSD, YOLOv5 allows for real-time gesture recognition without compromising detection accuracy. Its modularity also enables customization and fine-tuning for specific sign language variants, such as ASL (American Sign Language) or ISL (Indian Sign Language).

The system is designed to work in diverse lighting and background conditions, enhancing its robustness in real-world environments. To achieve this, the dataset includes gesture samples taken under varying angles, lighting scenarios, and hand orientations. Data augmentation techniques such as rotation, scaling, and flipping are applied during training to increase the model's generalizability and reduce overfitting.

Post-processing steps are used to refine detection results and eliminate false positives. Once a gesture is recognized, it is translated into text, which can optionally be converted to speech using a Text-to-Speech (TTS) engine. This multi-modal output ensures accessibility for both visual and auditory users.

To improve usability, a user-friendly interface is developed using Python (with libraries such as OpenCV, PyTorch, and Tkinter) or integrated into a mobile application using platforms like Android Studio. The interface displays live video, recognized signs, and the resulting translations in real-time. Future versions of the system can include support for continuous gesture recognition, dynamic signs (involving motion), and full sentence translations using Natural Language Processing (NLP).

The main objectives of this project are:

- To develop a gesture recognition model using YOLOv5.
- To create a dataset or use an existing one containing labeled sign gestures.
- To perform model training, testing, and performance evaluation.
- To implement real-time video processing for gesture detection.
- To build a functional interface that outputs translated text and/or speech.

The system's performance is evaluated using metrics such as accuracy, precision, recall, and frames per second (FPS) during inference. Testing on unseen gestures ensures the model's reliability and accuracy in varied settings.

The potential applications of this project extend beyond personal communication. It can be used in education, healthcare, customer service, and public kiosks to assist in conversations between hearing-impaired individuals and others. With further development, this system could evolve into a complete sign language interpreter, supporting complex grammar and continuous sign recognition.

This project highlights the potential of combining deep learning and computer vision to create socially impactful technology. By leveraging YOLOv5's capabilities, the Sign Language Translator becomes a tool not only for recognition but also for empowerment—giving voice to those who communicate with their hands.

II. LITERATURE REVIEW

The development of sign language recognition systems has gained momentum in recent years due to the increasing need for inclusive communication tools. Various approaches have been proposed, each utilizing different machine learning and computer vision techniques to enhance the accuracy and efficiency of gesture recognition.

1. Vision-Based Gesture Recognition Systems:

Earlier gesture recognition systems relied on traditional computer vision techniques such as color segmentation, background subtraction, and edge detection. However, these methods often struggled under varying lighting conditions and complex backgrounds. To overcome these challenges, deep learning models, particularly Convolutional Neural Networks (CNNs), have become the preferred approach due to their robustness and accuracy in image classification tasks.

2. Use of YOLO for Real-Time Detection:

The YOLO (You Only Look Once) object detection algorithm has revolutionized real-time detection tasks. The YOLOv5 model, in particular, has demonstrated significant improvements in speed and precision, making it ideal for detecting dynamic hand gestures. Research such as Bochkovskiy et al. (2020) on YOLOv4 laid the foundation for YOLOv5, which has been widely adopted in gesture-based applications for its balance between inference speed and detection accuracy.

3. Integration of CNNs with YOLO Models:

Several studies have explored the integration of CNNs with object detection algorithms to improve classification accuracy. For example, Mehta et al. (2018) developed a deep CNN-based system for American Sign Language (ASL) recognition, achieving high accuracy on static images. Combining YOLOv5's localization abilities with CNN's classification strength enables efficient recognition of complex and dynamic gestures in real time.

4. Real-World Sign Language Translation Systems:

Recent works have focused on real-time sign language translation systems that convert gestures into text or speech. Kaur and Kaur (2020) proposed a system for converting Indian Sign Language to text using CNNs and achieved promising results. Similarly, Kumar et al. (2021) utilized transfer learning with deep learning models for multilingual sign language recognition, highlighting the need for scalability in diverse linguistic environments.

5. Facial Recognition and Attendance Systems:

Parallel to gesture recognition, facial recognition technologies have been extensively researched, particularly for security and attendance systems. Algorithms such as Haar Cascade Classifiers and LBPH (Local Binary Patterns Histograms) have been implemented in real-time systems using OpenCV, demonstrating reliable performance under various conditions (Viola and Jones, 2001). These methods laid the groundwork for applying computer vision in real-world accessibility solutions.

III. PROPOSED SYSTEM

The proposed methodology utilizes the YOLOv5 model and Convolutional Neural Networks (CNNs) for realtime sign language translation. The process begins with the Data Collection phase, where a dataset of diverse sign language gestures is captured using video or image recordings. This data is then preprocessed, including steps like resizing, normalization, and labeling, to standardize the input for model training.

In the Model Training phase, YOLOv5 is employed for rapid hand gesture detection, while CNNs are used to classify gestures based on shape, position, and movement. The combined model learns to identify unique features of each sign language gesture accurately. Once trained, the model is stored and ready for real-time applications.

After training, the system moves to the Real-Time Translation phase. A camera captures live hand gestures, and YOLOv5 processes these images to detect hands in real time. The CNN then classifies the detected gestures, converting them into the

corresponding text or speech output. This enables seamless communication for sign language users without manual input.

Finally, the Communication and Data Management phase ensures that all translations are accurately recorded and stored. The system continuously refines its performance based on user feedback, and the user interface displays the translated gestures clearly. This real-time sign language recognition system ensures accessible communication, promoting inclusivity for the deaf and mute community.

The methodology ensures a scalable and efficient sign language translation system. Using YOLOv5 for Realtime detection and CNNs for classification, it operates with minimal hardware, adaptable across environments enhancing accessibility and inclusivity in diverse settings.

IV. SYSTEM ARCHITECTURE AND DESIGN

The Sign Language Translation System employs a modular architecture designed for real-time communication and accessibility. It begins with a Data Collection module, capturing diverse sign language gestures through video recordings or images.

The Data Preprocessing module standardizes the collected data by resizing and labeling the images for accurate training. The Model Training module utilizes YOLOv5 for rapid gesture detection and CNNs for detailed gesture classification, storing the trained model for future use.

Once trained, the system's Prediction module monitors gestures in real-time, detecting and translating them into text or speech using the combined YOLOv5 and CNN capabilities. The Backend processes gesture recognition, translating hand movements with low latency, and storing the data for continuous improvement.

This architecture ensures a scalable and efficient solution, applicable in various environments to facilitate seamless communication for deaf and mute individuals, fostering inclusivity and accessibility.

V. RESULTS AND DISCUSSION

The implementation of the Sign Language Translation System involved several stages, beginning with the development of a streamlined interface using Python's Tkinter for ease of use. The system leverages OpenCV for image capture, with YOLOv5 handling real-time hand gesture detection. Collected gesture data was processed and stored in a database, enabling accurate training using Convolutional Neural Networks (CNNs) for gesture classification. During the translation phase, the system captures live video footage, detects gestures with YOLOv5, and classifies them through CNNs to convert them into text or speech. The system then displays the translated output, ensuring seamless communication.

The results of the system's implementation were encouraging, achieving an accuracy rate of around 88% for gesture recognition under well-lit conditions. The real-time detection capabilities provided smooth communication with low latency. The system effectively recognized a wide range of gestures, translating them in under a second per gesture.

Additionally, it handled varying hand angles and lighting conditions, maintaining consistent performance. While accuracy could be further improved with more advanced models, YOLOv5 and CNNs provided a balanced solution suitable for environments with limited computational resources.

The system's performance met expectations for small to medium-scale applications, such as educational or workplace settings, offering a reliable, AI-driven communication tool for the deaf and mute community.

VI. CONCLUSION AND FUTURE SCOPE

The Sign Language Translation System provides a reliable and accessible solution for real-time communication between sign language users and nonsigners. Utilizing YOLOv5 for quick gesture detection and CNNs for precise classification, the system effectively bridges communication gaps. Its adaptability, minimal hardware requirements, and

real-time performance make it suitable for various environments, including educational and professional settings.

Future enhancements will focus on supporting additional sign languages and improving recognition accuracy under diverse conditions. Integrating advanced models and natural language processing could further enhance contextual understanding. The project highlights the potential of AI-driven technologies in promoting inclusivity and accessibility.

While the current implementation focuses on static sign recognition (e.g., letters and numbers), there are several areas for enhancement and expansion:

Extend the model to support dynamic gestures (gestures involving motion), enabling the recognition of full words and sentences using video sequence analysis or LSTM-based models.

Adapt the system to support various sign languages like ASL (American Sign Language), BSL (British Sign Language), and ISL (Indian Sign Language), allowing for broader application.

Use Natural Language Processing to form grammatically correct sentences from sequences of recognized signs, improving communication clarity. Develop lightweight versions of the system for Android, iOS, and web platforms, making the technology more accessible to users everywhere.

Incorporate AR to display translations in real time through smart glasses or AR apps for a more immersive experience.

Implement reverse translation where spoken words are converted into animated sign gestures, enabling two-way communication.

Optimize the model for edge computing devices using techniques like quantization and pruning to reduce power consumption and latency.

Embed the translator into customer service kiosks, public transportation terminals, classrooms, and hospitals to aid hearing-impaired users in various scenarios.

VII. REFERENCES

[1] Patil, S., Shinde, S., & Deshmukh, P. (2017). Realtime face detection and attendance management system using OpenCV. International Journal of Computer Applications, 162(6), 1-5.

[2] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, I-511.

[3] Gupta, S., & Pandey, V. (2018). A hybrid approachfor attendance system using face recognition and biometric data. International Journal of Computer Science and Information Technologies, 9(2), 85-90.

[4] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 815-823.

[5] Wang, Z., Li, P., & Yang, Y. (2020). Face recognitionbased attendance system using deep learning. Journal of Artificial Intelligence and Neural Networks, 29(2), 1320.