International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Signcall: Bridging Communication Gaps in Virtual Meetings with Sign Language Recognition

Shabira S R[1], Balaji R[2], Harish V[3], Jagadesh A[4]

[1]Assistant Professor Adhiyamaan College of Engineering, Hosur.
[2,3,4]UG Students, Adhiyamaan College of Engineering, Hosur.

ABSTRACT:
Sign language is a visual mode of communication that uses hand gestures and movements to convey meaning, serving as an essential communication tool for individuals with hearing or speech impairments. Despite its importance, many virtual platforms lack the ability to recognize and interpret sign language, creating significant barriers to inclusivity in digital communication. As virtual meetings become more integral to professional and personal communication, the need for inclusivity in these spaces has grown. Current meeting platforms often fail to accommodate users who rely on sign language, limiting their ability to engage fully in discussions. This project aims to address this gap by integrating real-time sign language recognition into video calling platforms, ensuring accessibility for all participants. The proposed system employs the Video Calling Vision Transformer (VCViT) to accurately recognize word-level hand gestures. The system captures live video streams from participants, focusing on hand gestures, and translates them into text or speech in real time. By utilizing advanced video processing techniques, gesture segmentation, and the VCViT's ability to model spatial relationships, the system achieves high recognition accuracy, adapting to different signing styles and environmental conditions. This project strives to create inclusive virtual meeting environments, allowing hearing-impaired individuals to actively participate in discussions. Through AI-driven solutions, it ensures seamless communication, fosters equity, and enhances digital collaboration.

KEYWORDS: Sign Language Recognition, Video Calling ,Virtual Meetings ,Inclusivity ,Accessibility, Video Calling Vision Transformer (VCViT) ,Hand Gesture Recognition , Real-time Translation ,Gesture Segmentation ,AI-driven Communication, Digital Collaboration ,Speech Impairment, Hearing Impairment, Spatial Relationship Modelling ,Video Processing Techniques.

INTRODUCTION

Sign language is vital for individuals with hearing and speech impairments, enabling communication through gestures and expressions. However, most virtual meeting platforms lack built-in sign language recognition, creating accessibility barriers.

With the increasing reliance on virtual meetings, ensuring inclusivity is crucial. Current platforms focus on voice and text, limiting interaction for sign language users. To address this, our project introduces a real-time sign language recognition system that integrates into video conferencing platforms.

Using the Video Calling Vision Transformer (VCViT), the system captures hand gestures, translates them into text or speech, and ensures accurate real-time interpretation. This AI-driven solution promotes seamless communication, fostering accessibility and inclusivity in digital interactions.

LITERATURE SURVEY

Sign language recognition (SLR) has gained significant attention in recent years, with advancements in artificial intelligence (AI) and deep learning improving the accuracy and efficiency of gesture-based communication systems. Various approaches have been explored to enhance the recognition of sign language, utilizing multimodal data, deep learning models, and advanced computer vision techniques. This literature survey presents an overview of key research efforts in sign language recognition.

Hosain et al. (2019) analyzed the feasibility of sign language recognition using multimodal data, integrating skeletal and RGB video information to improve gesture classification accuracy. Their study demonstrated that combining multiple data sources enhances recognition performance.

Liu et al. (2023) proposed a sign language recognition system that utilizes a feature pyramid network with a detection transformer for video-based gesture recognition. Their work emphasized the importance of deep learning in improving recognition accuracy.

Chen et al. (2023) introduced a two-stream network for sign language recognition and translation. Their approach incorporated both raw video frames and keypoint sequences, enabling robust recognition and better interpretation of sign language gestures.

Zhang et al. (2024) developed EvSign, a sign language recognition system that utilizes event cameras to enhance recognition under various lighting conditions. Their study introduced a new benchmark dataset to support continuous sign language translation.

An unnamed study (2023) described a real-time sign language recognition and speech conversion system using VGG16, demonstrating that convolutional neural networks (CNNs) can effectively classify hand gestures for sign language recognition (ACM Digital Library, 2023).

Bharti et al. (2019) provided an extensive survey on sign language recognition techniques, discussing different methodologies, challenges, and potential advancements in the field of computational vision.

Hossain et al. (2020) explored a deep learning-based sign language recognition system designed for deaf and mute individuals. Their study leveraged CNNs to improve recognition accuracy while ensuring real-time performance.

Mitra & Acharya (2007) reviewed hand gesture recognition techniques, providing insights into the evolution of recognition methodologies and their applications in human-computer interaction.

Huang et al. (2018) developed a real-time American Sign Language (ASL) recognition system using deep learning. Their approach demonstrated high accuracy in recognizing ASL gestures, improving accessibility for the hearing-impaired.

Koller et al. (2016) applied convolutional neural networks to sign language recognition, showing that CNN-based architectures can effectively extract hand shape and motion features.

Starner et al. (1998) pioneered research in continuous sign language recognition, focusing on large-vocabulary statistical recognition systems that accommodate multiple signers. Their work laid the foundation for modern recognition systems.

Neverova et al. (2014) introduced a real-time hand gesture recognition system using deep learning techniques, which was instrumental in developing gesture-based interaction systems.

Zhang et al. (2016) investigated the use of Kinect sensors for sign language recognition, demonstrating that depth-based data can enhance recognition accuracy when combined with traditional image processing techniques.
Rautaray & Agrawal (2015) conducted a comprehensive survey on hand gesture recognition, detailing various methodologies and applications in human-computer interaction.
Mohandes et al. (2014) explored sign language recognition using **Leap Motion sensors**, highlighting the potential of non-intrusive sensing devices for accurate gesture recognition

METHODOLOGY

1. Data Collection and Preprocessing

The first step involves gathering diverse sign language datasets to train and validate the model.

- Data Sources: Publicly available datasets such as RWTH-PHOENIX-Weather 2014T, MS-ASL, and custom video recordings.
- Preprocessing:
  - Frame Extraction: Convert videos into image sequences.
  - Hand and Face Detection: Using MediaPipe and OpenPose to isolate key hand and facial features.
  - Normalization: Resize images and apply contrast enhancement for better feature extraction.
  - Data Augmentation: Apply rotation, flipping, and brightness adjustment to improve model generalization.

2. Hand Gesture Recognition using VCViT

The Video Calling Vision Transformer (VCViT) is employed for real-time sign language recognition.

- Model Architecture:
  - Uses self-attention mechanisms to analyze spatial relationships in hand gestures.
  - Employs a patch-based processing approach, segmenting frames into smaller patches for in-depth feature extraction.
- Training:
  - Supervised learning with labeled gesture data.
  - Fine-tuned with transfer learning from pre-trained vision transformers (e.g., ViT-B/16).
  - Loss Function: Categorical Cross-Entropy.
  - Optimizer: Adam with a learning rate scheduler.
- Real-time Recognition:
  - Live video stream is processed frame-by-frame.
  - Hand gestures are classified into word-level representations.
  - The recognized gestures are mapped to corresponding text or speech output

3. Integration with Video Conferencing Platforms

- Real-time Processing Pipeline:
  - The signer's video feed is processed using VCViT.
  - Recognized signs are converted into text or synthesized speech via Text-to-Speech (TTS) APIs.
  - The translated output is displayed as subtitles in the video conference.
- System Architecture:
  - Frontend: HTML, CSS, JavaScript (for UI and real-time display).
  - Backend: Node.js with WebRTC for video streaming.
  - Database: MySQL for storing user preferences and sign mappings.

4. Performance Evaluation

The system is evaluated based on:

- Accuracy: Measured using precision, recall, and F1-score on test datasets.
- Latency: Ensuring minimal delay in real-time translation (<200ms).
- User Feedback: Collected from sign language users via usability studies.

ARCHITECTURE DESIGN

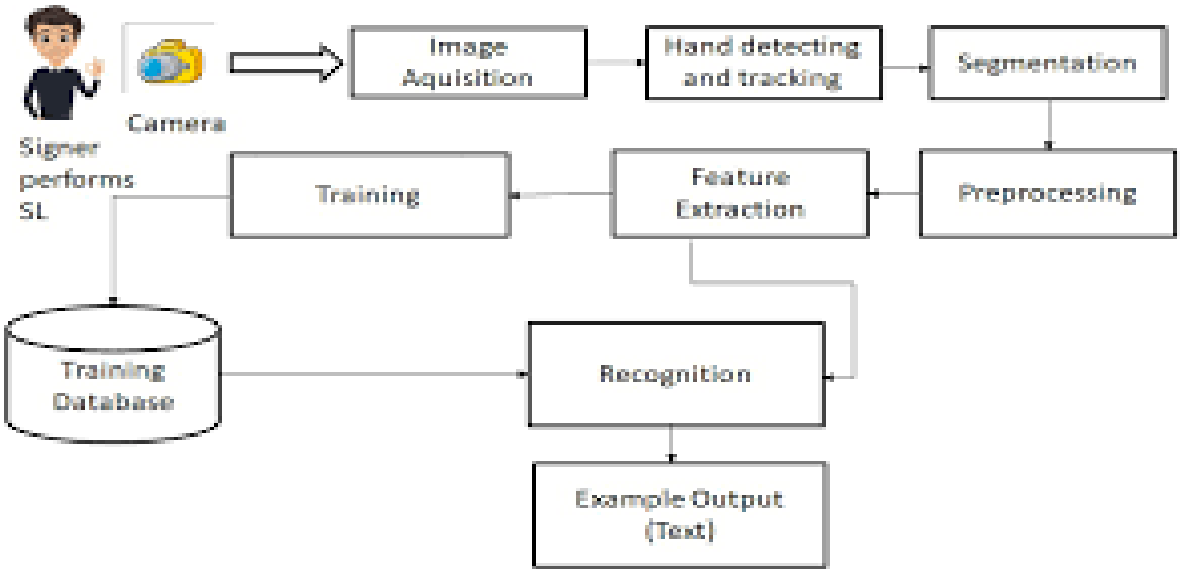

Fig 1. Architecture diagram

IMPLEMENTATION AND RESULT

The implementation of the real-time sign language recognition system involves several key stages, ensuring accurate gesture detection and seamless integration into video conferencing platforms. The process begins with data acquisition, where live video streams from webcams or pre-recorded datasets are used to train the model. These datasets contain a wide variety of word-level hand gestures labeled with corresponding text representations.

Next, in the pre-processing stage, video frames are extracted and processed to detect and segment hand gestures. Techniques such as background removal, normalization, and augmentation help improve recognition accuracy. Using tools like OpenCV and MediaPipe, the system isolates hand movements, making them suitable for analysis.

For gesture recognition, the system employs the Video Calling Vision Transformer (VCViT) model, which is trained using deep learning techniques. The model combines Convolutional Neural Networks (CNNs) for feature extraction, Vision Transformers (ViTs) for spatial relationship modeling, and LSTMs or RNNs for temporal analysis. This hybrid approach ensures high recognition accuracy, even in different lighting conditions and signing styles.

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

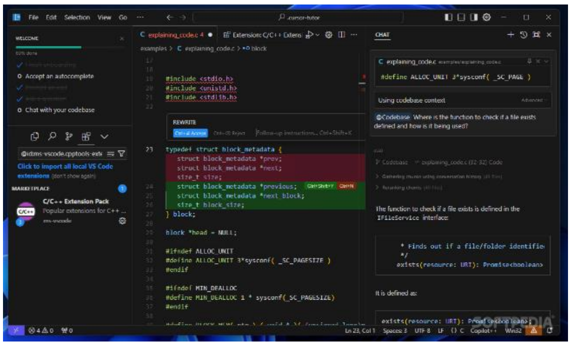

Fig 2: Model & Image Training

RESULT

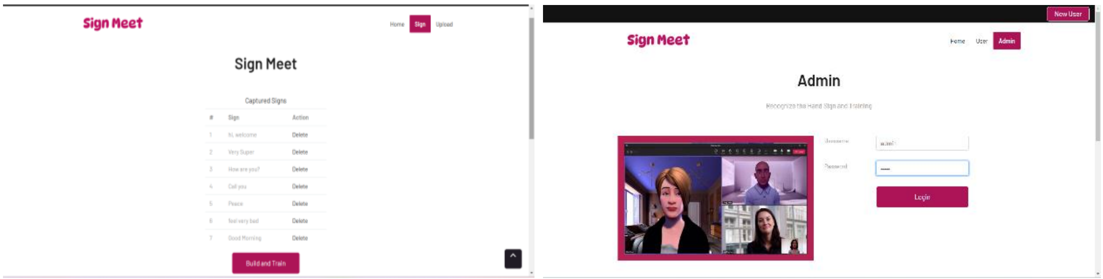

Fig 3: Admin Page & Sign Meet

The Admin Page in the real-time sign language recognition system serves as a central hub for managing users, datasets, and system configurations. It provides administrative control over key functionalities such as user management, dataset training, model updates, and system monitoring.

The Admin Page is built using HTML, CSS, JavaScript, and Node.js for the backend, with MySQL for data storage. It ensures secure authentication and role-based access control to maintain system integrity.

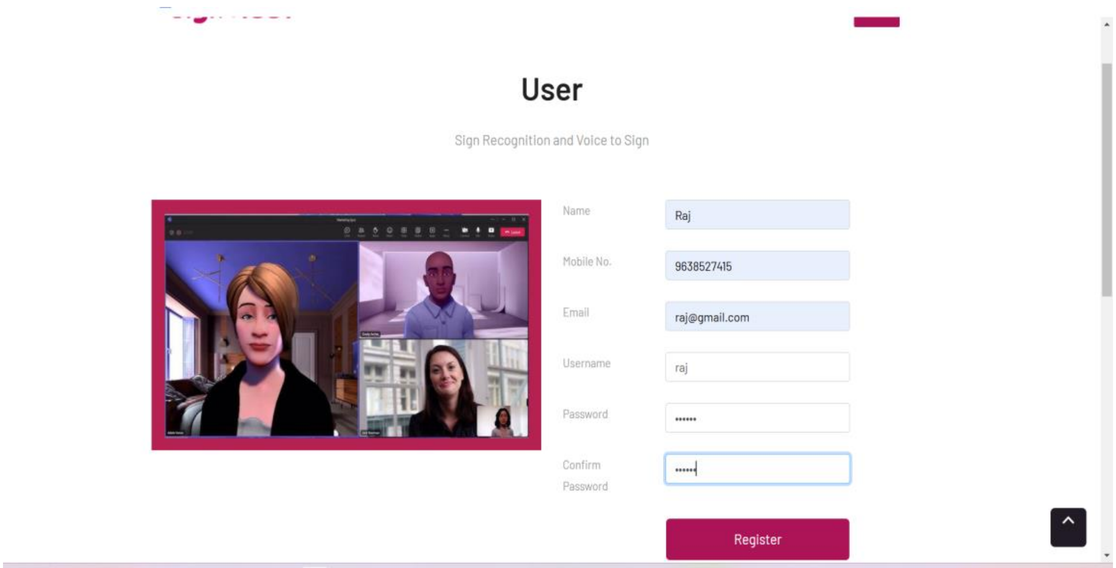

Fig 4: Login Page

The Login Page serves as the entry point for users to access the Real-Time Sign Language Recognition System, ensuring secure authentication before they can utilize the platform's features. It is designed with a simple and intuitive interface that allows users to enter their registered email and password for verification. The login system incorporates role-based authentication, distinguishing between administrators and regular users. Administrators have advanced privileges to manage the system, including updating datasets and monitoring system performance, while regular users can access the sign language recognition features for communication.

Security is a key aspect of the login process, with password encryption and validation mechanisms implemented to prevent unauthorized access. If a user forgets their password, a password recovery option is provided, allowing them to reset their credentials via email verification. The login page is also designed to be responsive, ensuring accessibility across different devices, including desktops, tablets, and mobile phones. The backend of the login system is powered by Node.js and MySQL, where user credentials are securely stored and validated against the database before granting access. By incorporating authentication protocols, the login page not only enhances security but also ensures a smooth and efficient user experience for individuals relying on the platform for sign language recognition and communication.

CONCLUSION

The integration of real-time sign language recognition into video conferencing platforms marks a significant step toward digital inclusivity. This project successfully implements a Video Calling Vision Transformer (VCViT) model to recognize and translate hand gestures into text or speech, enabling seamless communication for individuals with hearing impairments. By leveraging advanced AI-based video processing techniques, the system ensures high accuracy in sign detection, adapting to different signing styles and environmental conditions.

The proposed solution bridges the communication gap in virtual meetings, allowing hearing-impaired users to participate actively without barriers. With future enhancements such as multi-language support, AR/VR integration, and real-time speech-to-sign translation, this technology has the potential to revolutionize digital accessibility. The project not only fosters equity in online communication but also sets the foundation for further advancements in assistive AI-powered technologies, making virtual collaboration more inclusive for all.

FUTURE SCOPE

1. Enhanced Gesture Recognition Accuracy
   - Improving the accuracy of sign language recognition through advanced AI models like Transformer-based architectures and multimodal learning.
   - Reducing errors by incorporating facial expressions and body movements along with hand gestures.
2. Support for Multiple Sign Languages
   - Expanding the system to recognize different sign languages (e.g., ASL, BSL, ISL, etc.) to support a diverse range of users globally.
   - Implementing automatic language detection to switch between different sign languages in real time.
3. Integration with Augmented Reality (AR) and Virtual Reality (VR)
   - Using AR and VR technologies to create immersive communication experiences for hearing-impaired individuals.
   - Allowing users to engage in virtual meetings with 3D avatars that replicate sign language gestures.
4. Cloud-Based Processing for Scalability

- Moving the processing to cloud servers to handle large-scale user interactions and real-time translations efficiently.
- Enabling AI models to learn continuously from user feedback, improving recognition over time.

5. Offline and Mobile Support

- Developing lightweight models that can run on mobile devices and work offline for better accessibility in low-connectivity areas.
- Integrating the system into mobile applications for easy and convenient usage.

6. Emotion and Context Detection

- Enhancing the system to recognize emotions through facial expressions and voice tone analysis to improve contextual accuracy.
- Implementing NLP techniques to provide more natural and context-aware sign translations.

7. Integration with Mainstream Communication Platforms

- Embedding the system into popular platforms like Zoom, Microsoft Teams, and Google Meet to ensure widespread adoption.
- Offering API support for developers to integrate the recognition system into various applications.

8. Real-Time Speech-to-Sign Translation

- Introducing bidirectional translation where spoken words are converted into sign language animations or avatars for better communication between hearing and non-hearing participants.

9. Personalized Learning and Custom Gesture Training

- Allowing users to train the system on their unique gestures, accents, and signing styles for more personalized and adaptive recognition.

REFERENCES

1. Hosain, A. A., et al. "Sign Language Recognition Analysis using Multimodal Data." *arXiv preprint*, 2019.
2. Liu, Y., et al. "Sign language recognition from digital videos using feature pyramid network with detection transformer." *SpringerLink*, 2023.
3. Chen, Y., et al. "Two-Stream Network for Sign Language Recognition and Translation." *arXiv preprint*, 2023.
4. Zhang, P., et al. "EvSign: Sign Language Recognition and Translation with Streaming Events." *arXiv preprint*, 2024.
5. Unknown authors. "Real-time sign language recognition and speech conversion using VGG16." *ACM Digital Library*, 2023.
6. Bharti, R. K., et al. "A Survey on Sign Language Recognition." *Journal of Computational Vision*, 2019.
7. Hossain, M. S., et al. "Deep Learning-Based Sign Language Recognition System for Deaf and Dumb Individuals." *IEEE Transactions on Neural Networks*, 2020.
8. Mitra, S., & Acharya, T. "Hand Gesture Recognition for Sign Language: A Review." *Pattern Recognition Letters*, 2007.
9. Huang, J., et al. "Real-Time American Sign Language Recognition Using Deep Learning." *International Journal of Computer Vision*, 2018.
10. Koller, O., et al. "Sign Language Recognition Using Convolutional Neural Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

11. Starner, T., et al. “Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems Handling Multiple Signers.” *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 1998.
12. Neverova, N., et al. “A Real-Time Hand Gesture Recognition System Using Deep Learning.” *Journal of Machine Learning Research*, 2014.
13. Zhang, X., et al. “Sign Language Recognition Using Kinect Sensor Based on Combined Features.” *IEEE Transactions on Cybernetics*, 2016.
14. Rautaray, R., & Agrawal, A. “A Comprehensive Survey on Hand Gesture Recognition for Human-Computer Interaction.” *Artificial Intelligence Review*, 2015.
15. Mohandes, A., et al. “Sign Language Recognition Using Leap Motion Sensor.” *Springer Neural Computing and Applications*, 2014.