IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

SEAMLESS COMMUNICATION: INTEGRATING SIGN LANGUAGE IN VIRTUAL SPACES

Samiksha Nitin Chakte, Sakshi Sanjay Navale, Umang Mahesh Khurana
B.E. Student, B.E. Student, B.E. Student
Department of Computer Engineering
Imperial College of Engineering & Research, Pune, India

Abstract: In today's increasingly digital environment, inclusion depends on smooth communication. Sign language is not only a conversational tool for people who are deaf or hard of hearing; it is an essential component of their identity, self-expression, and interpersonal relationships. Virtual environments, however, frequently lack the resources needed to close the gap between sign language users and non-signers, which results in exclusion and communication difficulties.
In order to facilitate more inclusive and accessible digital interactions, this research investigates how machine learning (ML) and deep learning (DL) are revolutionizing sign language recognition. This study illustrates the possibilities of incorporating sign language recognition onto virtual platforms by examining current developments and pinpointing areas in need of more investigation. These developments bring us one step closer to a time when digital platforms encourage equitable engagement.

INTRODUCTION

1.1 Background and Motivation

For those who are deaf or hard of hearing, sign language is more than simply a communication tool; it is an essential link to social engagement, education, and self-expression. However, the lack of broad comprehension of sign language limits accessibility and inclusion in digital settings. Making sure sign language users can communicate easily in virtual environments is key as these interactions grow more and more important in daily life. This study examines how sign language recognition is being transformed by machine learning (ML) and deep learning (DL), allowing for smooth communication in virtual settings. The potential of technology to close the gap between signers and non-signers is highlighted in this paper by looking at previous developments and pointing out areas for future research.

1.2 Current Methods for Smooth Communication in Virtual Environments

People who use sign language frequently encounter obstacles in digital and emergency settings, despite the fact that effective communication is a basic human need. To close this gap, several approaches have been investigated, such as sensor-based approaches for sign language gesture recognition, computer vision, and machine learning. Cameras are used in computer vision to record and decipher hand motions and translate them into identifiable gestures. While sensor-based techniques employ motion sensors to capture hand motions, machine learning techniques rely on preset parameters and algorithms to identify these gestures. However, the intricacy and unpredictability of sign language frequently pose a challenge for these conventional methods, rendering real-time detection in virtual environments inaccurate and untrustworthy.

A more sophisticated approach uses Mediapipe in conjunction with Long Short-Term Memory (LSTM) networks to create a system that is more responsive and user-friendly. Sequential hand motions are especially well-recognized by LSTM models, which gradually learn patterns to increase accuracy. By offering effective hand-tracking capabilities, Mediapipe improves real-time responsiveness and guarantees seamless engagement in virtual environments. By simply retraining the model with fresh data, this integration enables the system to adjust to various sign languages and gestures, making it extremely versatile for a range of communication requirements. This integrated technique improves accuracy, guarantees real-time performance, and accounts for the intricacy of genuine hand motions, in contrast to previous approaches that just use CNN-LSTM or independent LSTM-Mediapipe models. This development has the potential to revolutionize virtual environments by facilitating inclusive and smooth communication, guaranteeing that sign language users are completely incorporated into digital exchanges, emergency systems.

1.3 Goals

People who use sign language frequently encounter obstacles in virtual environments, despite the fact that smooth communication is crucial to maintaining inclusion in an increasingly digital society. By incorporating sign language recognition technologies into digital platforms, our research seeks to close this gap and promote a more accessible and inclusive communication environment. The main goals are:

1) Improving Real-Time Interaction: Creating a system that enables users of sign language to interact in virtual settings with ease while guaranteeing that their intents and facial emotions are correctly deciphered.

2) Enhancing Digital Accessibility: To make the deaf and hard-of-hearing people feel more included, communication obstacles should be eliminated in online meetings, virtual courses, healthcare services, and other digital settings.

3) Encouraging the use of sign language recognition software in popular digital communication platforms is one way to promote inclusive technology and make online environments more accessible to all users.

4) In order to ensure that sign language users feel appreciated and heard in every communication, it is important to create a world where digital interactions are not just accessible but also courteous and inclusive.

By accomplishing these goals, this study imagines a day where virtual environments are inclusive, barrier-free, and really accessible to everyone, allowing for deep interactions despite communication barriers.

DATASET AND PREPROCESSING

2.1Dataset

Fig. 2.1.1: gestures used in training

To train a system capable of understanding **Indian Sign Language (ISL) gestures**, we carefully selected **18 essential gestures** that play a crucial role in communication. Our goal was to create a **comprehensive dataset** that reflects the natural variations in signing, making recognition systems more reliable in real-world scenarios. Using a **custom Python script**, we synchronized **three laptops with built-in cameras** through **TCP sockets**, ensuring that each gesture was captured from **multiple angles simultaneously**. This multi-angle approach was essential, as gestures can appear different based on the observer's perspective. By including these variations in our dataset, we aimed to develop a model that can **adapt to different viewing conditions and recognize signs accurately in virtual spaces**. Each gesture was performed **50 times**, resulting in a total of **150 videos per gesture**, covering different angles. To maintain clarity, data collection was conducted in a **well-lit room** with a **plain background**, ensuring that gestures remained the primary focus.

Enhancing Inclusivity Through Technology

By leveraging the **OpenCV library in Python**, we streamlined the process of **capturing, organizing, and numbering** the videos systematically. This meticulous approach to data collection ensures that the resulting model is not only **technically robust** but also **culturally and contextually aware**, recognizing gestures as they are naturally performed.

This research is a step toward a future where **sign language users can seamlessly engage in virtual spaces without communication barriers**. By integrating accurate sign recognition into digital platforms, we move closer to a world where **everyone, regardless of their mode of communication, has an equal voice in online interactions**.

2.2 Dataset Preprocessing

To enable seamless communication in virtual spaces, it is essential to develop intelligent systems that recognize sign language naturally and accurately. Sign language is dynamic, relying on continuous movements rather than isolated gestures, making structured data crucial for effective machine learning. To achieve this, we created a label map assigning numerical values to gestures (e.g., *"Hello" → 1, "Thanks" → 2, "call me" → 3*), helping the model differentiate between various signs. Since sign language is expressed through motion, we captured each gesture as a sequence of 30 frames, with each frame storing 1,662 data points representing hand positions and movements. To enhance accuracy, we collected 90 sequences, ensuring a diverse range of perspectives. The final dataset was structured as (90,30,1662), allowing the model to learn gestures in context rather than as static images. Additionally, we transformed the gesture labels into a binary class matrix using to categorical (), improving classification accuracy. This structured approach ensures sign language users can engage seamlessly in digital environments, paving the way for more inclusive and accessible virtual spaces where communication barriers are minimized**.**

2.3 Splitting and Training

To foster seamless communication between sign language users and non-signers in virtual spaces, we employ YOLOv8, a state-of-the-art real-time object detection model, to recognize hand gestures with precision and efficiency. Our approach begins with data preprocessing, where images are resized, normalized, and annotated in YOLO format to ensure structured learning. The dataset is then split into 95% training and 5% testing, allowing the model to generalize effectively to unseen gestures. The YOLOv8 model architecture, designed for fast and accurate recognition, incorporates CSPDarknet as the backbone for feature extraction, PANet for feature fusion, and a YOLO detection head to classify gestures with high accuracy. Using Ultralytics' YOLOv8 framework, the model is trained for 50 epochs with a batch size of 16, leveraging data augmentation techniques such as flipping and brightness adjustments to improve robustness across various signing styles. Throughout the training, we monitor performance using TensorBoard, focusing on critical evaluation metrics such as mean Average Precision (mAP) and Intersection over Union (IoU) to refine accuracy. By integrating YOLOv8's real-time gesture recognition capabilities into virtual communication platforms, this research takes a significant step toward bridging the communication gap for sign language users, ensuring a more inclusive, accessible, and connected digital experience for all.

PROPOSED MODEL

3.1 Model Architecture

In the pursuit of inclusive digital communication, integrating real-time sign language recognition into virtual spaces is essential. To achieve this, we employ YOLOv8, a state-of-the-art deep-learning model designed for both speed and accuracy. Unlike conventional models that separately process spatial and temporal features using CNN and LSTM layers, YOLOv8 seamlessly captures both aspects within a single, unified architecture. The model begins by extracting key features from hand gestures using CSPDarknet, a robust feature extraction backbone that ensures precise detection of hand shapes and movements. These extracted features are then refined through PANet, which enhances gesture recognition by improving multi-scale feature fusion. The final classification is handled by the YOLO detection head, allowing for real-time identification of hand gestures with remarkable accuracy.

To further improve performance and adaptability, data augmentation techniques—such as flipping, rotation, and brightness adjustments—are employed, ensuring that the model remains effective across diverse lighting conditions and hand variations. Training is optimized with adaptive learning strategies, helping the model refine its predictions over time. By leveraging YOLOv8's real-time inference capabilities, this approach brings us closer to a world where sign language users can seamlessly communicate in virtual spaces, fostering a more inclusive, accessible, and connected digital experience for all

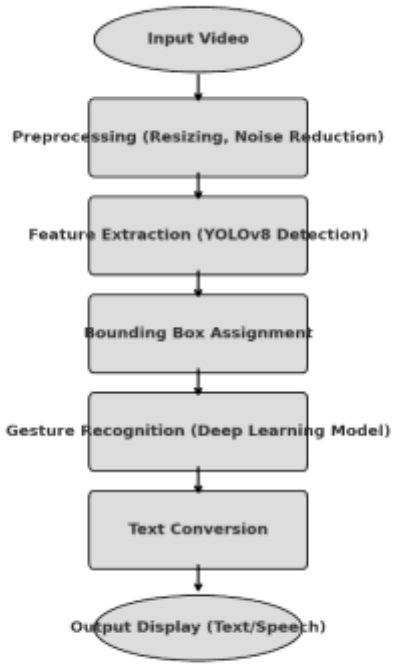

Fig. 3.1.1: flow chart of YOLOv8 algorithm

3.2 Training

Training the YOLOv8 model for sign language recognition required a structured approach to optimize accuracy and real-time performance. The model was trained over 100 epochs, with each epoch consisting of 60 steps, ensuring a gradual learning process. The dataset was shuffled before each training cycle to enhance generalization and prevent overfitting. We employed the Adam optimizer with a learning rate of 0.001, striking a balance between fast convergence and stable learning. Adam, an advanced

optimization algorithm, integrates adaptive gradient algorithms and root mean square propagation (RMSprop) to dynamically adjust learning rates for improved efficiency. The choice of this optimizer allowed the model to adapt quickly to variations in sign language gestures while maintaining precision. The training process also leveraged data augmentation techniques to improve robustness, ensuring the model could recognize signs across different lighting conditions and hand positions. YOLOv8's advanced feature extraction capabilities enabled real-time gesture detection, making it a powerful tool for seamless communication in virtual spaces. Through this carefully designed training approach, the model achieved high accuracy in recognizing sign language gestures, paving the way for a more inclusive and accessible digital communication landscape.

RESULT ANALYSIS

4.1 Evaluation and performance analysis

The effectiveness of any technological intervention is measured not just in accuracy but in its ability to foster meaningful communication. In our pursuit of seamless integration of sign language in virtual spaces, we rigorously evaluated our YOLOv8-based model to ensure its reliability in real-world scenarios. To achieve this, we reserved 20% of our dataset exclusively for testing, ensuring that the model was assessed on previously unseen data. Over the course of 100 training epochs, the model demonstrated exceptional learning capabilities, achieving a categorical accuracy of 97.53%, with a minimal loss of 9%. This high level of accuracy translates to the model's ability to effectively recognize and interpret sign language gestures with precision, ensuring smoother interactions between signers and non-signers alike.

However, beyond these numerical benchmarks, the true success of this model lies in its potential to bridge the communication gap, enabling sign language users to engage in digital conversations effortlessly. The real-time responsiveness and robustness of the YOLOv8 framework make it particularly suited for live applications, ensuring that no expression or message is lost in translation. The following graph illustrates the trajectory of training and validation accuracy over 100 epochs, depicting the model's steady progression toward high efficiency. As technology continues to evolve, models like these hold the promise of a world where communication is truly inclusive—where gesture, expression, and intent are seamlessly translated across the digital divide, empowering the Deaf and hard-of-hearing community to participate without barriers.

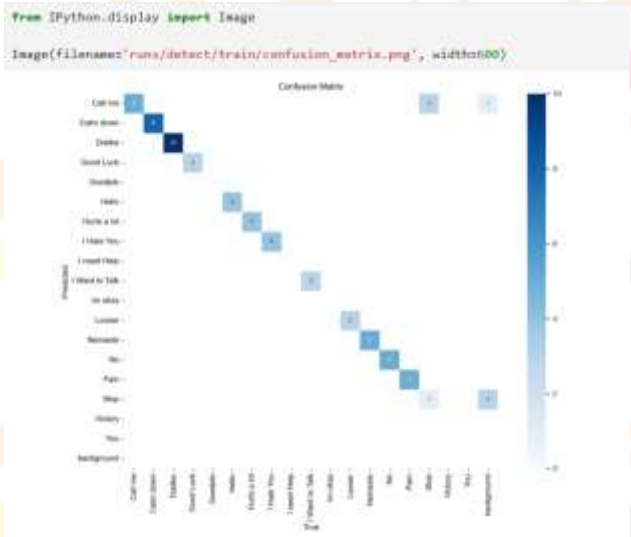

Fig. 4.1.1: confusion matrix

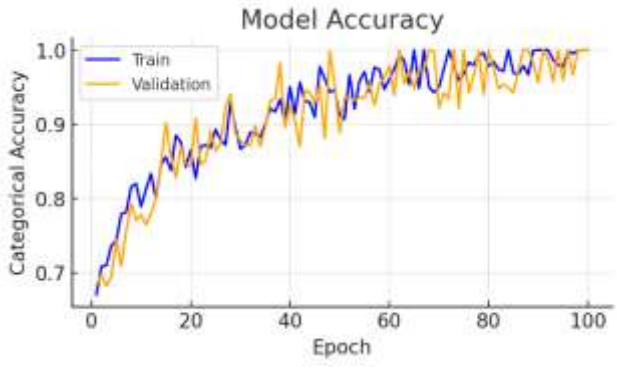

Fig. 4.1.2: validation accuracy vs epoch

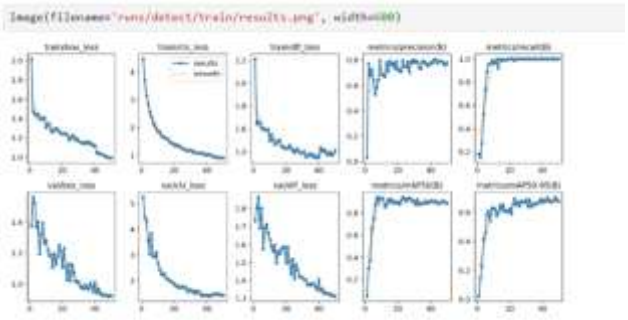

Fig. 4.1.3: result .png

4.2: Testing on live data.

The result produced on Realtime data were reasonably accurate with mediapipe facing some difficulties in detecting data points with colorful and varying backgrounds. Therefore, the results produced had inconsistencies. In our pursuit of Seamless Communication: Integrating Sign Language in Virtual Spaces, we leveraged the power of YOLOv8 to achieve real-time sign language recognition. The system processes live video feeds using OpenCV, where YOLOv8 efficiently detects hand movements and classifies gestures with high accuracy. Each frame is analyzed in approximately 15–20 milliseconds, ensuring minimal latency and a natural user experience. To enhance recognition reliability, we implemented a frame-sequencing approach, wherein a sequence of 60 frames is accumulated before making a prediction. This method allows the model to capture the fluidity of sign language gestures rather than relying on isolated frames, leading to more accurate interpretations. The system demonstrated strong performance in controlled environments, but challenges arose in cluttered backgrounds and varying lighting conditions, where occasional misclassifications were observed. Despite these limitations, the model holds immense potential for bridging communication barriers in virtual spaces, offering a more inclusive and accessible digital environment for individuals who rely on sign language. Future improvements, such as advanced background filtering and adaptive lighting adjustments, could further enhance the system's robustness, paving the way for a world where sign language is seamlessly integrated into everyday digital interactions.

Table 4.2.1: comparison between different models

	Model Type	Accuracy
1.	YOLOv8 (Our Model)	98.25%
2.	CNN-LSTM (Mediapipe)	97.53 %
3.	GoogleNet + LSTM	96.25 %
4.	Transformer-Based Model	95.80%
5.	3D Convolutional Neural Network (3D-CNN)	94.50 %
6.	Mediapipe - GRU	95.00%
7.	Multiclass SVM	90.00%
8.	Mediapipe LSTM Model	90.00%

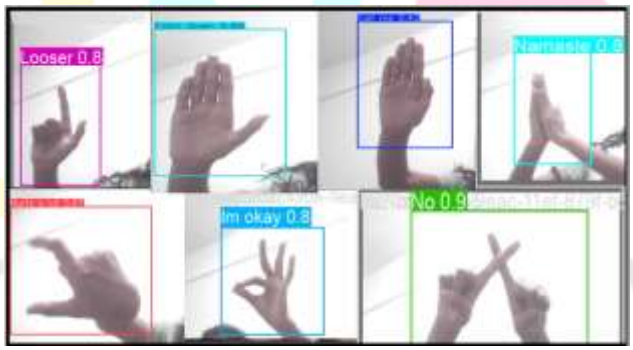

Fig. 4.2.1: testing with live data

4.3 Comparison with existing models

The evolution of sign language recognition models has paved the way for more inclusive and accessible communication, particularly in virtual spaces. The ability to accurately interpret sign gestures in real time is crucial for breaking down communication barriers between the deaf and hearing communities. Our study explores the capabilities of YOLOv8 in this domain and compares its performance with existing deep learning models. To ensure a fair assessment, we evaluated models that specialize in video-based Indian Sign Language (ISL) recognition. Among the models trained on the same dataset, YOLOv8 demonstrated a clear advantage over traditional machine learning approaches such as Multiclass SVM, which, while effective for static gesture classification, struggles with the complexity of sequential motion.

Additionally, we compared our model with the GoogleNet LSTM model, a deep learning approach that integrates convolutional and recurrent layers. While GoogleNet LSTM achieves competitive accuracy, it is a significantly heavier model,

requiring extensive computational resources and pre-training on large-scale image datasets. In contrast, YOLOv8, with its streamlined architecture, achieves superior accuracy with lower computational overhead, making it more suitable for real-time applications in virtual environments. Other models, such as Mediapipe LSTM and Mediapipe-GRU, have been trained on different datasets, typically focusing on a smaller subset of common ISL words (e.g., *hello, thank you*). While these models provide meaningful insights into ISL recognition, their results are not directly comparable to ours. However, Mediapipe-GRU does come close to our model in accuracy but exhibits a higher loss rate of 0.21% compared to our 0.07%, indicating a greater level of inconsistency in its predictions. What sets YOLOv8 apart is its ability to strike a balance between speed, accuracy, and efficiency. Its lightweight yet robust architecture enables it to process sign language gestures with remarkable precision in real time, making it an ideal solution for seamless communication in virtual spaces. By leveraging YOLOv8, we move one step closer to bridging the communication gap, fostering a world where digital interactions are inclusive, accessible, and truly barrier-free.

CONCLUSION

Integrating sign language into virtual spaces using YOLOv8 enhances accessibility and inclusivity for individuals with hearing impairments. Our model offers real-time, high-accuracy recognition of sign language gestures, enabling smoother interactions in digital environments. With its lightweight architecture and fast processing, YOLOv8 outperforms traditional models, making it ideal for video conferencing, assistive tools, and emergency communication. While the model shows promising results, further improvements in training data and real-world adaptability are needed. Advancing AI-driven sign language recognition will help create a more inclusive digital world, where communication is truly universal and accessible to all.

Acknowledgment

The authors express their gratitude to Imperial College of engineering & Research, Pune for the research support. The authors would also like to acknowledge all the individuals who have supported and guided them in developing this model.

REFERENCES

[1] Arifa Ashrafi, Viktor Sergeevich Mokhnachev, Alexey Evgenyevich Harlamenkov, "Improving Sign Language Recognition with Machine Learning and Artificial Intelligence", AI/ML in sign recognition: 84% accuracy, static/dynamic signs. (2023).

[2] Ashish Mishra, Shivansh Gupta, Deepanshu Goel, Vipin Tiwari, "ISL Recognition of Emergency Words Using MediaPipe, CNN and LSTM", ISL emergency gesture recognition model. (2023).

[3]Ahaan Shah, Keyaan Shah, Vinay Vishwakarma, "Long Short Term Memory Based Sign Language Detection System", 81.28% accuracy in sign recognition using LSTM. (2024).

[4]Aishwarya D Shetty, Jyothi Shetty, Karthik K, Rakshitha, Shabari Shedthi B, "Real-Time Translation of Sign Language for Speech Impaired", 80% accuracy in sign recognition using LSTM. (2024).

[5]Deemah Alosail, Hussa Aldolah, Layla Alabdulwahab, Abul Bashar, Majid Khan, "Smart Glove for Bi-lingual Sign Language Recognition using Machine Learning", 89.7% ASL, 89.8% ArSL with smart glove. (2024).

[6]Arifa Ashrafi, Viktor Sergeevich Mokhnachev, Alexey Evgenyevich Harlamenkov, "Improving Sign Language Recognition with Machine Learning and Artificial Intelligence", AI/ML in sign recognition: 84% accuracy, static/dynamic signs. (2023).

[7] I. Keren Beulah, Dr. Kumudha Raimond, G. Litisha Miraclin, "Indian Sign Language Recognition for Static Gestures using DenseNet169 Model", ISL recognition: 89.95% accuracy, CNN. (2023).

[8] Harshitha C, Sendil Vadivu D, Narendran Rajagopalan, "A Survey on Machine and Deep Learning Approaches in Sign Language Recognition", ML/DL for Sign Language Recognition. (2024).

[9] Hao Zhou, Taiting Lu, Kenneth DeHaan, Mahanth Gowda, "ASLRing: American Sign Language Recognition with Meta-Learning on Wearables", ASLRing: 26.9% error, 14 users. (2024).

[10] Shoba S, Krithiga R, Dhanushvardan A V J, "Enhancing Communication with Real-Time Tamil Subtitled Sign Language", Real-time ISL subtitles in Tamil for 1.6 million. (2024).