International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Hands of Hope: A Real-Time Sign Language Recognition System Using Deep Learning

Mr. Mandar Gawande[1],Ms. Pooja Maknor[2],Mr. Shubham Haral[3],Ms. Vaishanvi Nerpgar[4]

Matoshri College of Engineering and Research Center Nashik

ABSTRACT—

In a world where communication is a fundamental human right, the "Hands of Hope" aims to bridge the gap between the hearing and deaf communities through advanced sign language detection. This paper presents a novel approach to real-time sign language recognition by leveraging deep learning and computer vision techniques. By utilizing a robust dataset of sign language gestures, our system accurately interprets and translates sign language into text and speech. The proposed solution incorporates state-of-the-art convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to enhance gesture recognition precision and efficiency. Experimental results demonstrate significant improvements in detection accuracy and processing speed compared to existing methods. Our work not only advances the field of assistive This paper details the implementation of the "Hands of Hope" system, focusing on the methodologies, data preprocessing, and the integration of CNN and RNN architectures. Additionally, the work highlights the challenges encountered during development and the performance of the system in real-world scenarios. By enabling real-time translation of sign language into text and speech, this project aims to improve accessibility and foster inclusivity for individuals with hearing impairments.

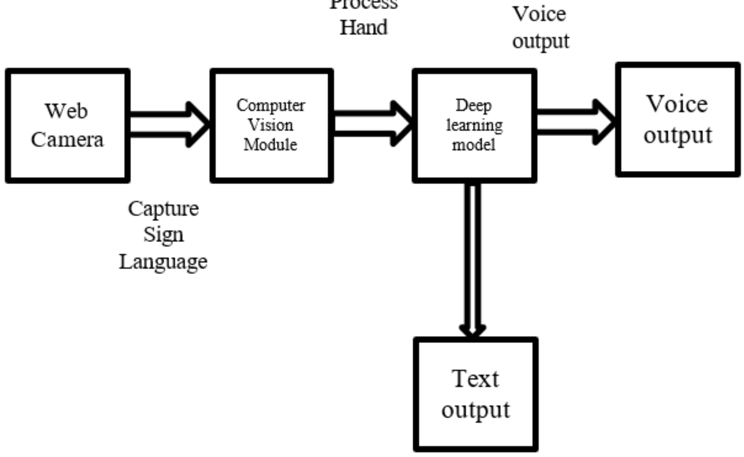
Keywords—Convolutional Neural Networks (CNNs) ,Recurrent Neural Networks (RNNs),Real-time Processing, Sign language.

I .Introduction

Communication is a fundamental human right and an essential aspect of building inclusive societies. However, for millions of individuals with hearing impairments, effective communication remains a persistent challenge due to the limited understanding and adoption of sign language in mainstream interactions. Sign language serves as a vital medium of expression for the deaf community, yet the lack of real-time translation solutions significantly hinders their ability to engage seamlessly with others [1]. The ―Hands of Hope‖ project seeks to address this challenge by introducing an seamless system for real-time sign language recognition. This project integrates includes technologies in deep learning and computer vision to bridge the communication gap between the hearing and deaf communities.[2] By leveraging robust datasets of sign language gestures, the system translates these visual expression into text and speech, enabling dynamic and effective interactions. The approach utilizes convolutional neural network(CNNs) for feature extraction and recurrent neural networks(RNNs) for sequence learning, ensuring high accuracy and efficiency. In In addition to its technical underpinnings, the "Hands of Hope" system emphasizes the importance of creating a robust dataset of sign language gestures. This dataset is carefully curated to represent diverse signs, ensuring the system can cater to various regional and cultural variations of sign language. Data preprocessing techniques such as noise reduction, normalization, and augmentation are employed to enhance the dataset's quality and ensure consistency during training. Furthermore, the system is designed with scalability in mind, making it adaptable to include additional languages and gestures in the future

II .Proposed Model

Fig:1 Workflow of Real-Time Sign Language Recognition System.



Above Fig. 1 illustrates the workflow of the real-time sign language recognition system, which includes the web camera for gesture capture, the computer vision module for hand detection and preprocessing, the deep learning model for gesture classification, and the output modules for displaying text and generating voice predictions technologies but also fosters inclusivity, enabling smoother interactions for individuals with hearing impairments. The Hands of Hope project envisions a future where communication barriers are dismantled, fostering greater understanding and empathy across diverse populations

Web Camera

The first step of the process begins with a **web camera**, which serves as the input device, capturing real-time video of the user's hand gestures.

- **Input Role:** The web camera plays a crucial role by streaming continuous frames of video, which are converted into digital images. These images serve as the primary input to the system, allowing for the capture of hand movements.
- **Data Capture:** The camera is designed to support dynamic gesture capture, meaning it can adjust to variations in lighting, user positioning, and the speed of gestures. This ensures the system remains responsive to different conditions during usage.
- **Resolution Considerations:** The quality of the camera's resolution is key to accurately detecting fine details of the hand gestures. A higher resolution ensures that smaller, intricate movements such as finger positions, angles, and orientations can be effectively captured and processed by the system.

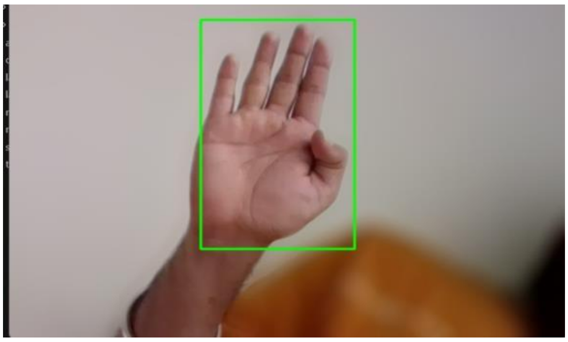


Fig:2 The Region of Interest (ROI)

To discard the unwanted area of the video frame, a particular area is fixed as a Region of Interest (ROI). Figure 2 shows the Region of Interest.

Computer Vision Module

This module processes the video feed to isolate and extract meaningful features related to hand gestures. Mediapipe is the primary tool used here.

Hand Detection

- **Hand Landmark Extraction**: Mediapipe identifies 21 key landmarks on the user's hand. These landmarks include important points like the fingertips, knuckles, and wrist.
  - **Mathematical Representation**: Each landmark is represented as a 3D coordinate (x,y,z(x, y, z)where:
    - x and y are normalized image coordinates.
    - z represents the depth (distance from the camera).
- **Bounding Box Generation**: The bounding box surrounding the hand is calculated using the minimum and maximum x, yx, yx, y coordinates from the detected landmarks.
- **Mathematical Formulation**: bbox_min[0]=min(x_i)−m, bbox_min[1]=min(y_i)−m
  bbox_max[0]=maz(x_i)+m, bbox_max[1]=max(y_i)+m
  Here, m is the margin added to ensure the hand is fully captured.

Preprocessing and YAML File Generation

- **Normalization**: The landmarks are normalized to fit within a consistent range, improving compatibility with the model's input requirements.
- **Feature Extraction**:
  The extracted landmarks(x, y, z)are converted into a 1D array.
- **YAML File Creation**: For each gesture captured, a corresponding YAML file is generated containing the structured landmark data. This ensures a systematic approach to storing and reusing hand gesture data for training and testing in subsequent stages.
- **Deep Learning Mode** The processed data is passed into a pre-trained **neural network model**, which classifies the gesture. The model consists of multiple layers trained to recognize patterns in the input.

Model Architecture

- **Input Layer**: Accepts the flattened landmark data (1D array).
- **Hidden Layers**: Two dense layers with 128 and 64 neurons, using the ReLU activation function: $f(x)=\max(0,x)$
  This activation introduces non-linearity, allowing the model to learn complex patterns.
- **Output Layer**: A dense layer with Softmax activation to produce a probability distribution over n classes:

$$P(y_i|x)=\frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

Here, z_i represents the raw output for class i

Training Process

- **Loss Function**: Sparse Categorical Cross-Entropy is used:

$$L=-\sum^{n} y_i \log(\hat{y}_i)$$

Where y_i is the true label, and $\hat{y}_i$ is the predicted probability for class i.

- **Optimization**: The Adam optimizer updates model weights during training:

$$\theta=\theta-\eta\cdot\nabla_\theta J(\theta)$$

Where θ represents the model parameters, η is the learning rate, and J is the loss function.

Inference

- **Prediction**: For a given input, the model generates a class label by identifying the class with the highest probability: Predicted Class$=\arg\max_i P(y_i|x)$

Output Modules

The final classification result is mapped to a human-readable label and displayed in two forms:

Text Output

- The predicted gesture is displayed on the screen as text, making it easy for users to interpret the result.
- **Mapping**: Labels are stored in a JSON file, which maps integer class IDs to their corresponding gesture names.

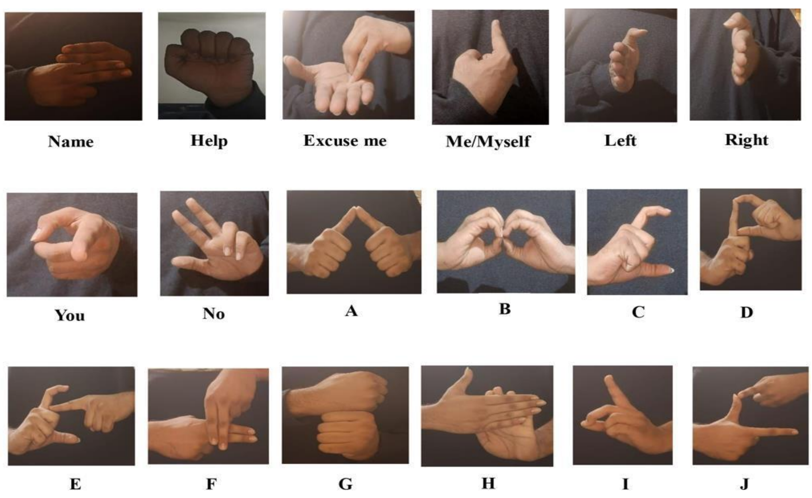
Voice Output

- Using Text-to-Speech (TTS), the system converts the predicted gesture into an audible format, enhancing usability for diverse audiences.
- **Dynamic Mapping**: The system dynamically retrieves labels for seamless integration between prediction and output.

EXPERIMENTAL SETUP AND EVALUATION

Single image of each character from the dataset for gesture recognition is shown.

Fig 3: Sign wise hand gesture classification



Each gesture was performed by multiple users under varying conditions to ensure diversity and robustness in the dataset. The system was trained and tested using this dataset, with an **80-20** train-test split

Training and Accuracy

The model was trained using a neural network architecture consisting of two dense hidden layers and a Softmax output layer. The training process spanned 100 epochs, with the Adam optimizer and Sparse Categorical Cross-Entropy loss function.

- **Training Accuracy**: The model achieved a high training accuracy of **96.83%**, indicating its ability to learn the patterns and features of the provided gestures effectively.

  The accuracies of individually sign are shown in Table I. The person wise results of our proposed model is shown in Table II

Table I. Sign wise Classification Accuracy

Signs	Accuracy	Average accuracy
Name	97%	
Help	99%	
Excuse me	96%	
Me	94%	
Left	97%	
Right	99%	
You	100%	
No	97%	96%
A	95%	
B	97%	
C	98%	
D	97%	
E	94%	
F	97%	
G	98%	
H	95%	
I	95%	

Table II. Person Wise Classification Accuracy

Person	Recognition Accuracy	Average Accuracy
P1	94.21%	
P2	92.33%	93.22%
P3	93.11%	

CONCLUSION

In this paper, have presented a real-time sign language recognition system utilizing CNN-RNN models for gesture recognition. The system effectively combines CNNs for feature extraction and RNNs for sequence learning, achieving high recognition accuracy. Frameworks like Mediapipe and TensorFlow further enhance the system’s performance and scalability. The proposed system outperforms traditional models in both efficiency and precision.

Future work will focus on expanding the system to include more diverse sign languages and dynamic gestures, catering to a broader user base. This project aims to bridge the communication gap between hearing and deaf communities, empowering individuals with hearing impairments to actively engage in socio-economic activities and fostering a more inclusive society

REFERENCE

1. M. R. Islam, U. K. Mitu, R. A. Bhuiyan, and J. Shin, ―Hand Gesture Feature Extraction Using Deep Convolutional Neural Network for Recognizing American Sign Language,‖ 2018 4th International Conference on Frontiers of Signal Processing (ICFSP), 2018, IEEE

2. Kishore, K. A. K. Yadav, and K. V. Karthikeyan, ―Online Hand Gesture Recognition and Classification for Deaf and Dumb,‖ 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), 2023, IEEE.

3. M. A. G. Alshehri, A. Alzahrani, and A. Alzahrani, ―Deep learning ap proaches for sign language recognition: A survey,‖ IEEE Access, vol. 11, pp. 15043–15055, 2023, doi: 10.1109/ACCESS.2023.3234569.

4. L. H. Nguyen, T. T. Nguyen, and D. T. Nguyen, ―Real-time sign language recognition using convolutional neural networks and LSTM,‖ Sensors, vol. 22, no. 4, p. 1505, 2022, doi: 10.3390/s22041505.
5. Y. Yang, Y. Zhou, and J. Zhang,―A novel framework for sign language recogni tion based on 3D hand tracking,‖ Journal of Ambient Intelligence and Human ized Computing, vol. 14, no. 3, pp. 1529–1538, 2023, doi: 10.1007/s12652-022
6. 04157-3.

7. K. F. O. E. S. B. S. V. S. R. P. S. D. K. A. M. K., ―Gesture recognition for sign language: A review of methods and applications,‖ Computers in Human Behavior, vol. 138, p. 107405, 2023, doi: 10.1016/j.chb.2023.107405.

   A. T. H. Wang and Y. Y. Wu, ―Integrating deep learning with image processing for robust sign language recognition,‖ Journal of Visual Communication and Image Representation, vol. 98, p. 103305, 2023, doi: 10.1016/j.jvci.2023.103305
8. S. M. Miah, M. A. M. Hasan, Y. Tomioka, and J. Shin, ―Hand Gesture Recognition for Multi-Culture Sign Language Using Graph and General Deep Learning Network,‖ IEEE Open Journal of the Computer Society, vol. 5, 2024, IEEE.
9. A. Al Abdullah, G. A. Amoudi, and H. S. Alghamdi, ―Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects,‖ IEEE Access, vol. 12, 2024, IEEE.
10. M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, T. S. Alrayes, H. Mathkour, and M. A. Mekhtiche, ―Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation,‖ IEEE Access, vol. 8, 2020, IEEE.