Smart Gestures-Sign Language Conversion

B. Srinivasulu[1], Ch. Jayanth[2], S. Deekshitha[2], S. Karthik[2]
[1]Associate Professor, [2]UG Scholar
Dept. of Information Technology, Vidya Jyothi Institute of Technology,
Hyderabad, Telangana, India.srinu@vjit.ac.in

***Abstract*—**The field of computer vision has long grappled with the challenge of sign language conversion. Despite numerous proposed solutions, none have achieved the portability required for implementation in standalone devices or applications. This study seeks to address this limitation by harnessing the potential of machine learning and recent deep learning advancements. Effective communication with hearing-impaired individuals remains a significant obstacle. Sign language has emerged as a crucial medium for those with hearing and speech disabilities to convey their thoughts and emotions to the world, facilitating smoother integration and reducing interpersonal complexities. However, the mere existence of sign language is insufficient to bridge the communication gap. The interpretation of sign gestures often proves challenging for those unfamiliar with the language or versed in different sign systems. Nonetheless, recent technological innovations offer the potential to narrow this longstanding communication divide through automated sign gesture detection. This research introduces a novel approach to sign language recognition, employing American Sign Language in conjunction with deep learning techniques

Keywords*—Block Chain; File Sharing; IPFS: Hyperledger Fabric:*

I. INTRODUCTION

Sign language serves as a crucial communication medium for individuals with hearing and speech impairments, enabling them to convey their thoughts, emotions, and ideas effectively. This visual language, comprising hand gestures, facial expressions, and body movements, facilitates rich and nuanced communication. However, the interpretation of sign gestures remains challenging for those unfamiliar with the language or versed in different sign systems, creating a significant barrier to effective interaction between hearing-impaired individuals and the broader community. [1,4]

This communication gap has long been a substantial obstacle in fostering inclusive environments and ensuring equal participation of hearing-impaired individuals in various aspects of society, including education, employment, and social interactions. The inability to bridge this linguistic divide often results in isolation, misunderstandings, and limited opportunities for those who rely on sign language as their primary mode of communication. [3]

The field of computer vision has made numerous attempts to address the challenge of sign language conversion, recognizing the potential for technology to serve as an intermediary between sign language users and those who do not understand sign language. Researchers have explored various approaches, including image processing techniques, gesture recognition algorithms, and sensor-based systems. Despite these proposed solutions, none have achieved the portability and accuracy required for widespread implementation in standalone devices or applications that could be easily integrated into daily life. Recent technological innovations, particularly in the realms of machine learning and deep learning, offer promising avenues to bridge this longstanding communication divide through automated sign gesture detection.[2]

These advanced computational techniques have demonstrated remarkable capabilities in pattern recognition, image classification, and natural language processing, making them well-suited for the complex task of interpreting sign language. [5]

This study introduces a novel approach to sign language recognition, leveraging American Sign Language (ASL) in conjunction with cutting-edge deep learning techniques. ASL, widely used in North America, serves as an ideal foundation for this research due to its well-established structure and extensive user base. By harnessing the potential of advanced machine learning algorithms, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), this research aims to develop a more portable and efficient solution for sign language conversion. [7]

The proposed system utilizes computer vision techniques to capture and analyze sign language gestures in real-time. Deep learning models are trained on large datasets of ASL signs, enabling them to recognize and interpret a wide range of gestures accurately. The system's architecture is designed to be lightweight and efficient, making it suitable for implementation on mobile devices and other portable platforms. One of the key challenges addressed in this study is the ability to handle the dynamic nature of sign language, including variations in signing speed, style, and regional dialects. [6,8]

The deep learning models are trained to be robust and adaptable, capable of recognizing signs across different users and contexts. Additionally, the system incorporates natural language processing techniques to convert recognized signs into coherent written or spoken language, ensuring accurate and contextually appropriate translations.

The ultimate goal of this research is to facilitate smoother integration of hearing-impaired individuals into society and reduce interpersonal complexities arising from communication barriers. By providing a reliable and accessible tool for sign language interpretation, this technology has the potential to enhance educational experiences, improve workplace communication, and foster more inclusive social interactions. Furthermore, the implications of this research extend beyond the immediate benefits for the hearing-impaired community. [9]

The development of advanced sign language recognition systems could contribute to broader advancements in human-computer interaction, gestural interfaces, and non-verbal communication analysis. These technologies could find applications in diverse fields such as robotics, virtual reality, and assistive technologies for individuals with various disabilities. As this research progresses, it is crucial to consider ethical implications and involve the deaf and hard-of-hearing community in the development process. Their insights and feedback will be invaluable in ensuring that the technology meets real-world needs and respects the cultural and linguistic aspects of sign language. [10,11]

In conclusion, this study represents a significant step forward in leveraging cutting-edge technology to address a longstanding communication challenge. By combining the expressive power of American Sign Language with the analytical capabilities of deep learning, we aim to create a more connected and inclusive world where language barriers no longer impede effective communication.

II. RELATED WORK

.This research presents a comprehensive overview of diverse approaches and research studies pertaining to sign language recognition utilizing various technologies and algorithms. The field of sign language recognition has garnered significant attention in recent years, with researchers exploring innovative methods to bridge communication gaps between deaf and hearing individuals. The key points encompass:

Numerous studies employ image processing and machine learning techniques for sign language recognition. These techniques form the foundation of many recognition systems, enabling computers to interpret and understand visual sign language input. [12]

Prevalent algorithms include Convolutional Neural Networks (CNN), Linear Discriminant Analysis (LDA), Hidden Markov Models (HMM), and deep learning models. CNNs have shown particular promise in image recognition tasks, while HMMs are effective in capturing the temporal aspects of sign language gestures. Deep learning models, with their ability to learn complex patterns, have pushed the boundaries of recognition accuracy.[14]

Recognition systems frequently convert signs into textual and/or auditory output. This conversion is crucial for enabling real-time communication between signers and non-signers, making these systems valuable tools for accessibility and inclusion. [13]

Datasets utilized in these studies comprise images and videos of letters, numerals, words, and sentences in various sign languages (e.g., Indian, Arabic, Pakistani). The diversity of these datasets reflects the global nature of sign language research and the need for systems that can accommodate different sign languages and dialects. [15]

Certain approaches utilize specialized hardware such as sensor-equipped gloves, while others rely on camera input and image processing. Sensor-based approaches offer high precision but may be less practical for everyday use, while camera-based systems are more accessible but face challenges in capturing subtle hand movements. [16]

Multiple studies focus on real-time recognition and mobile applications. The emphasis on real-time processing and mobile platforms highlights the practical applications of this research, aiming to provide immediate, on-the-go communication support for sign language users. [18]

Various preprocessing techniques and libraries (e.g., OpenCV) are employed for image processing. These tools help in enhancing image quality, segmenting relevant parts of the image, and preparing the data for subsequent analysis by recognition algorithms. [17]

Performance evaluation methodologies vary, with some studies utilizing limited datasets and others employing larger, more diverse datasets. The variation in evaluation methods underscores the need for standardized benchmarks to compare different approaches effectively.

Some research explores multi-lingual sign language recognition and skeleton-aware approaches. Multi-lingual recognition systems aim to bridge communication gaps across different sign languages, while skeleton-aware approaches focus on capturing the underlying structure of human gestures for more robust recognition. [19]

Additionally, researchers are exploring the integration of contextual information and natural language processing techniques to improve the accuracy and naturalness of sign language translation. Some studies are also investigating the use of augmented reality and virtual reality technologies to enhance sign language learning and recognition systems.

The field continues to evolve, with ongoing challenges including handling variations in signing styles, addressing occlusions and complex backgrounds in real-world scenarios, and developing systems that can recognize continuous sign language in natural conversations. Future research directions may include the development of more sophisticated multi-modal systems that combine visual, depth, and motion data for improved recognition accuracy. [20,21]

III. PROPOSED METHODOLOGY

The system employs a vision-based approach. All signs are represented using bare hands, thus eliminating the need for artificial devices in the interaction process.

Data Set Generation

For the project, we attempted to locate pre-existing datasets; however, we were unable to find datasets in the form of raw images that met our requirements. The only available datasets were in the form of RGB values. Consequently, we decided to create our own dataset. The steps we followed to create our dataset are as follows. We utilized the Open Computer Vision (OpenCV) library to produce our dataset.

Initially, we captured approximately 400 images of each symbol in American Sign Language (ASL) for training purposes and approximately 100 images per symbol for testing purposes.

Gesture Classification:

Algorithm Layer 1:
1. Apply a Gaussian Blur filter and threshold to the frame captured with OpenCV to obtain the processed image after feature extraction.
2. This processed image is then passed to the Convolutional Neural Network (CNN) model for prediction. If a letter is detected for more than 50 frames, it is printed and considered for word formation.
3. Space between words is denoted using the blank symbol.

Algorithm Layer 2:
1. We detect various sets of symbols that yield similar results upon detection.
2. We then classify these sets using classifiers specifically designed for those sets.

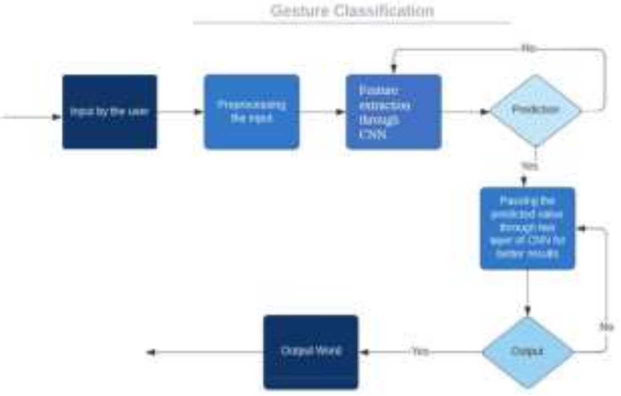

Gesture Classification
Input by the user
Preprocessing the input
Feature extraction through CNN
Prediction
No
Yes
Passing the predicted value through two layer of CNN for better results
Output
No
Yes
Output Word

Fig2: Gesture Classification

Convolutional neural networks for feature learning and classification

1. Initial Convolution Layer: The process begins with an input image of 128x128 pixels. This image undergoes initial processing in the first convolutional layer, which employs 32 filter weights (each 3x3 pixels). The outcome is a set of 126x126 pixel images, one corresponding to each filter weight.

2. Primary Pooling Layer: These images are then subjected to downsampling via max pooling with a 2x2 window. This operation retains the maximum value within each 2x2 array, effectively reducing the image dimensions to 63x63 pixels.

3. Secondary Convolution Layer: The 63x63 pixel outputs from the preceding pooling layer are then fed into the second convolutional layer. This layer also utilizes 32 filter weights (3x3 pixels each), producing 60x60 pixel images as a result.

4. Secondary Pooling Layer: Another round of down sampling occurs, again using 2x2 max pooling, further reducing the image resolution to 30x30 pixels.

5. First Fully Connected Layer: The resulting images serve as input to a fully connected layer comprising 128 neurons. The output from the second convolutional layer undergoes reshaping into an array of 30x30x32 = 28800 values. This array of 28800 values constitutes the input to this layer. The layer's output is subsequently directed to the Second Fully Connected Layer. To combat overfitting, a dropout layer with a rate of 0.5 is implemented.

6. Second Fully Connected Layer: The output generated by the First Fully Connected Layer is subsequently processed by another fully connected layer containing 96 neurons.

7. Output Layer: The final layer receives input from the Second Fully Connected Layer. The number of neurons in this layer corresponds to the number of classes in the classification task (alphabets plus a blank symbol).

IV. IMPLEMENTATION

Feature Extraction and Representation:

Images are encoded as three-dimensional matrices, where the dimensions correspond to the image's height and width, while the depth represents pixel values (1 for grayscale, 3 for RGB).

These pixel values serve as input for CNN to derive significant features.

Convolutional Neural Network (Cnn):

In contrast to standard Neural Networks, CNN layers organize neurons in three dimensions: width, height, and depth. Neurons in each layer connect to a limited area (window size) of the previous layer, rather than forming full connections. The final output layer's dimensions match the number of classes, as the CNN structure ultimately compresses the entire image into a single vector of class scores.

Convolution Layer:

This layer utilizes a small window (typically 5*5) that extends through the input matrix's depth. It contains trainable filters of the window size. The window moves by a stride (usually 1), calculating the dot product of filter components and input values at each location. This procedure generates a 2D activation matrix, showing the filter's response across various spatial positions. The network develops filters that activate when detecting particular visual elements, such as edges or color patterns.

Pooling Layer:

Pooling layers diminish the activation matrix size and trainable parameters. Two types exist: a. Max Pooling: Employing a window (e.g., 2*2), it chooses the maximum of 4 values. This process continues, yielding an activation matrix half the initial size. b. Average Pooling: This method takes into account all values within a window.

Fully Connected Layer:

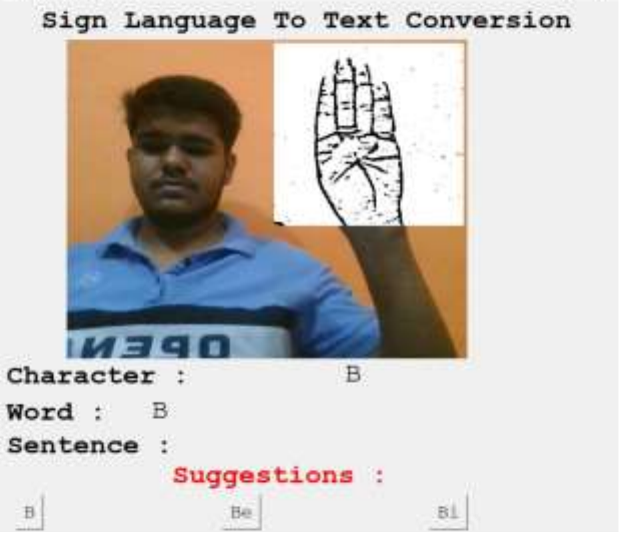
Unlike the convolution layer where neurons connect to a local region, the fully connected layer links all inputs to neurons.

Final Output Layer:

The fully connected layer's output connects to the final neuron layer (with neuron count equal to the total class number), estimating the likelihood of an image belonging to various classes.

Contours play a crucial role in image processing for object detection and recognition. In our research, we have utilized contours to distinguish and identify hands from their surroundings. These contours are defined as curves that connect continuous points sharing the same color. The initial phase of contour detection in OpenCV is analogous to identifying white objects against a black backdrop, necessitating the application of Inverted Binary Thresholding. The subsequent phase involves delineating the contours, which can be employed to outline any shape when boundary points are known. Below Figures showcases various gestures from our recognition system, accompanied by their respective contours.

.



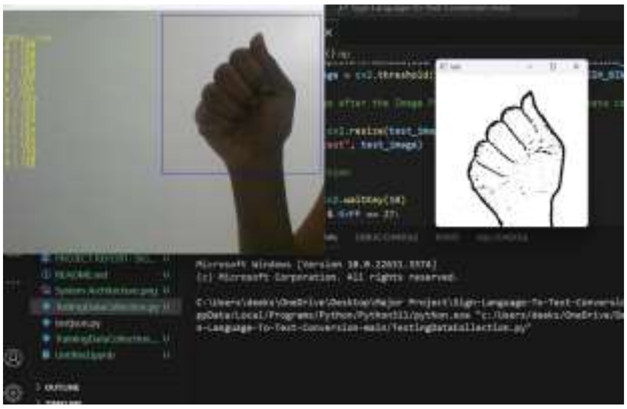
Sign Language To Text Conversion

Character : B

Word : B

Sentence :

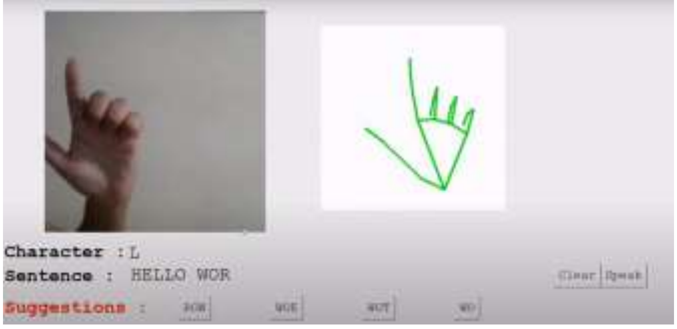
Fig: SWord Formation



```
Microsoft Windows [Version 10.0.22631.3374]
(c) Microsoft Corporation. All rights reserved.

C:\Users\deeks\OneDrive\Desktop\Major Project\Sign-Language-To-Text-Conversio
ppData/Local/Programs/Python/Python311/python.exe "c:/Users/deeks/OneDrive/D
n-Language-To-Text-Conversion-main/TestingDataCollection.py"
```

Fig Data Collection



Character : L

Sentence : HELLO WOR

Fig: sentence formation

V. CONCLUSION

This study presents a promising approach to address the longstanding challenge of sign language conversion in computer vision. By leveraging machine learning and deep learning advancements, particularly in American Sign Language recognition, the research aims to overcome the portability limitations of previous solutions. The proposed method has the potential to significantly improve communication between hearing-impaired individuals and those unfamiliar with sign language. As technology continues to evolve, this innovative approach may pave the way for more accessible and efficient automated sign gesture detection systems, ultimately fostering better integration and reducing communication barriers for the hearing-impaired community.

REFERENCES

[1] Choi SG, Park Y, Sohn CB (2022) Dataset transformation system for sign language recognition based on image classification network. Appl Sci 12(19):10075

[2] Bhatti Z, Muhammad F, Malik HAM, Hussain M, Chandio H, Channa S, Mahar Z (2021) Text to animation for sign language of Urdu and Sindhi. IKSP J Emerg Trends Basic Appl Sci 1(1):08–14

[3] Jiang S, Sun B, Wang L, Bai Y, Li K, Fu Y (2021) Sign language recognition via skeleton-aware multi-model ensemble. arXiv preprint arXiv:2110.06161.

[4] Shen X, Zheng Z, Yang Y (2022) Stepnet: spatial-temporal part-aware network for sign language recognition. arXiv preprint arXiv:2212.12857

[5] Ivanko D, Ryumin D, Karpov A (2019) Automatic lip-reading of hearing impaired people. Int Arch Photogramm Remote Sens Spat Inf Sci 42:97–101

[6] Kumar GA, William JH (2022) Development of visual-only speech recognition system for mute people. Circuits Syst Signal Process, pp 1–21

[7] Ryumina E, Ivanko D (2022) Emotional speech recognition based on lip-reading. In: International conference on speech and computer, pp 616–625. Springer International Publishing, Cham

[8] Zhou H, Zhou W, Zhou Y, Li H (2020) Spatial-temporal multi-cue network for continuous sign language recognition. Proc AAAI Conf Artif Intell 34(7):13009–13016

[9] Kratimenos A, Pavlakos G, Maragos P (2021) Independent sign language recognition with 3d body, hands, and face reconstruction. In: ICASSP 2021–2021 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 4270–4274, IEEE

[10] Cheripelli R, ChS, Appana DK. New Model to Store and Manage Private Healthcare Records Securely Using Block Chain Technologies. In: Bangabandhu and Digital Bangladesh. ICBBDB 2021. Commun. Comput. Inf. Sci. 2022, 1550.

[11] Kaur R, Kaswan S (2022) Conversion of Punjabi sign language using animation. In: Rising threats in expert applications and solutions: proceedings of FICR-TEAS 2022, pp 175–185, Springer Nature Singapore, Singapore

[12] Sumana M, Hegde SS, Wadawadagi SN, Sujana DV, Narasimhan VG (2022) Smart tutoring system for the specially challenged children. In: Society 5.0: smart future towards enhancing the quality of society, pp 113–130. Nature Singapore, Singapore

[13] Aasofwala N, Verma S, Patel K (2021) A novel speech to sign communication model for gujarati language. In: 2021 Third international conference on inventive research in computing applications (ICIRCA), pp 1–5, IEEE

[14] Amal H, Reny RA, Prathap BR (2020) Hand kinesics in indian sign language using NLP techniques with SVM based polarity. Int J Eng Adv Technol IJEAT 9(4):2249–8958

[15] Mouti S, Rihawi S (2020) IoT and sign language system (SLS). Int