International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

American Hand Gesture Recognition into Text Translation

Ankita Pingale
Department of information technology
ABMSP's Anantrao pawar college of engineering and research
Pune,Maharashtra
ankitapingale2002@gmail.com

Prajakta khaire
Department of information technology
ABMSP's Anantrao pawar college of engineering and research
Pune,Maharashtra.
prajakta.khaire@abmspcoerpune.org

Tanmay Phuse
Department of information technology
ABMSP's Anantrao pawar college of engineering and research
Pune,Maharashtra.
tanmay9623bari@gmailcom

Rohan Madane
Department of information technology
ABMSP's Anantrao pawar college of engineering and research
Pune,Maharashtra
rohanvmadane@gmail.com

Mangesh Markad
Department of information technology
ABMSP's Anantrao pawar college of engineering and research
Pune,Maharashtra
mamgeshmarkadcom@gmail.com

***Abstract*—**

American Sign Language (ASL) recognition represents a transformative advancement in technology, automating the interpretation of human expressions into comprehensible language constructs. This automation simplifies what has traditionally been a complex and laborintensive process, allowing for construction of systems tailored to specific linguistic and recognition tasks. ASL recognition systems leverage cuttingedge technologies to analyze and interpret gestures, expressions, and movements in real-time, revolutionizing the domain of human computer interaction and accessibility solutions. The merging of ASL recognition into modern technological frameworks has proven particularly effective in domains such as expression analysis, gesture-based communication, and language processing applications. By automating the recognition of ASL, the systems can enhance expressively tasks like expression detection, gesture typing in American English, and natural language translation. The architectures produced are not only excellent in accuracy but also save computational overhead, making them extremely efficient and suitable for real world applications.

Keywords— American Sign Language (ASL), Sign Language Recognition (SLR), Image Processing, Gesture Recognition, Deep Learning, Real-Time Recognition, Hand Gesture Detection, Human-Computer Interaction (HCI).

I. INTRODUCTION

Sign Language Recognition (commonly abbreviated as SLR) is a pivotal computational task that focuses on identifying and interpreting gestures from sign languages. This task is particularly crucial in today's digital era, as it seeks to bridge the communication divide experienced by individuals with hearing impairments. SLR employs advanced technologies, like image processing and natural language processing, to facilitate this process. In SLR systems, the user's hand movements are recorded by a camera, analyzed using sophisticated algorithm,and translated into English text or speech, conveying the user's intended message.

In addition to developing recognition systems, the project emphasizes creating a comprehensive ASL dashboard for real-time tracking, visualization, and evaluation of different architectures' performance. This dashboard will serve as a vital tool for researchers and developers, providing insights into the effectiveness of various models and enabling iterative improvements. By offering real-time feedback and performance metrics, the dashboard ensures that the ASL recognition models are continuously optimized to meet user demands and practical requirements

By applying advanced ASL recognition technologies, this project aspires to deliver mobile-ready solutions for speech and expression processing with improved accuracy, efficiency, and adaptability.

II. IMPORTANCE OF TECHNOLOGY

This innovation makes advanced machine learning applications accessible on mobile platforms, ensuring that high-precision recognition and processing capabilities are available to users regardless of the hardware's computational limitations. By tackling issues associated with to hardware constraints, such as limited processing power and memory, ASL recognition models make it possible to bring high-quality language recognition solutions to a broader audience. In contrast to spoken languages, sign languages rely heavily on visual elements such as hand shapes, movements, and facial expressions, making their translation a complex endeavor.

These systems are designed not merely for recognize gestures and expressions accurately but also to translate them into meaningful text or speech outputs, bridging the communication gap for individuals relying on ASL. The ultimate goal is to create an inclusive, accessible, and technologically advanced solution that empowers individuals, enhances communication, and drives the adoption of ASL recognition in everyday life.

Image processing plays a important role in SLR, enabling the analysis and detection of hand gestures and other visual elements. This field leverages a variety of programming languages, including C, Python, Java, and Matlab.

III. Literature review

Sharmila Gaikwad's study, "Recognition of American Sign Language Using Image Processing and Machine Learning," delves into gesture-based the capacity of gesture–based human Computer Interaction (HCI) systems. These systems facilitate non-contact interaction with computers through gestures, including those used in ASL. Gaikwad emphasizes the use of image processing together with machine learning to identify ASL gestures with efficiency. The study highlights the groundbreaking possibilities of such systems in bridging communication gaps for individuals who are deaf or have hearing impairments.

Yu Liu’s paper, "Sign Language Recognition from Digital Videos Using Feature Pyramid Network with Detection Transformer," introduces a novel framework that employs Feature Pyramid Networks (FPN) and Detection Transformers (DETR). This hybrid approach significantly enhances the system's ability to recognize static and dynamic ASL gestures. The paper emphasizes the importance of contextual information in ensuring accurate and robust ASL recognition in diverse real-world environments [2].

Teena Varma’s work, "Sign Language Detection Using Image Processing and Deep Learning," explores the translation of ASL gestures into American English text. Varma emphasizes the importance of preserving the cultural and semantic integrity of the original expressions while ensuring accurate translations. Combining visual processing with deep learning frameworks is targeted at enhancing detection accuracy while retaining the message intention [3].

Chenyang Zhang and Yingli Tian's work, "Multi-Modality American Sign Language Recognition," focuses using more than one modality, for instance, hand gestures, face expressions, and body poses, to enhance recognition of ASL. Using a dataset comprising RGB and depth video recordings, the authors demonstrate how multi-modal analysis enhances the system’s adaptability to varied signing styles and environmental conditions [4].

Parama Sridevi’s research, "Sign Language Recognition for Speech and Hearing Impaired by Image Processing in MATLAB," highlights the practical application of image processing techniques for ASL recognition. The proposed model operates effectively even in non-uniform background conditions, making it suitable for real-world applications. The study emphasizes the potential of such systems to enable interaction among people with and without speech and hearing disabilities [5].

Jinalee Jayeshkumar Raval's paper, "Real-Time Sign Language Recognition Using Computer Vision," explores the application of computer vision algorithms to facilitate real-time recognition of ASL. By pre-processing images to isolate the hands from the background, the study successfully recognizes the 24 letters of the English alphabet in ASL. The focus on real-time functionality ensures that the system is suitable for live communication scenarios [6].

Maryam Pahlevanzadeh's research, "Sign Language Recognition," primarily targets Taiwanese Sign Language (TSL) but offers insights applicable to ASL. The paper employs a two-layer classification system to enhance recognition accuracy by addressing variability in hand shapes and gestures. While focusing on TSL, the techniques discussed are adaptable to ASL, especially in handling contextual variations [7].

These studies collectively underscore the interdisciplinary nature of ASL recognition research. They highlight the role of image processing towards enhancing the usability and accuracy of recognition systems. By addressing challenges such as contextual understanding, real-time processing, and multi-modal analysis, these make way waysfor more inclusive communication technologies.

IV. RESEARCH METHODOLOGY

A. Paper Search

The framework for this paper was constructed using thorough studies on extensive research conducted across reputable academic databases and digital libraries.These sites were selected on account of offering full coverage for matters pertaining to computer vision and human-computer interaction. In order to make sure that all the relevant studies were identified, various search queries were carefully constructed using keywords such as "American Sign Language recognition," "gesture recognition," "image processing for ASL," "deep learning for ASL," "hand gesture recognition," and "computer vision for sign language." These keywords were selected to capture a broad spectrum of research focusing on ASL recognition using either image processing or machine learning approaches. Only studies directly addressing ASL recognition through these methodologies were shortlisted for further analysis.

B. Paper Screening Criteria

The choice of papers was based on a strict set of inclusion and exclusion criteria to determine the quality and applicability of the literature being reviewed. technical reports that directly addressed ASL recognition were considered. These papers either explored isolated ASL signs or continuous ASL recognition systems. Additionally, papers that presented experimental outcomes with clear performance metrics—such as classification accuracy, precision, recall, and real-time processing capabilities—were prioritized.

Special consideration was given to research that introduced innovative methodologies, novel datasets, or benchmarks for ASL recognition systems, as these contain key information on progress in the field. However, studies which did not aim explicitly at ASL (for instance, gesture recognition systems with general applicability rather than customized towards ASL) were ruled out. Non-peer-reviewed articles, white papers, and publications without scientific rigor or validation were also omitted from the review. Furthermore, works addressing isolated hand gestures rather than the complete linguistic structure of ASL were excluded, as the focus was on systems that treat ASL as a fully formed language and not as a set of discrete gestures.

C. Analysis of Paper Retrieval Results

After identifying and gathering the relevant literature, a systematic framework was applied to classify and analyze the findings. The studies were combined into two methodological categories: traditional image processing techniques and machine learning-based approaches. Traditional image processing methods—such as hand segmentation, background subtraction, edge detection, and contour analysis—were evaluated for their effectiveness in recognizing ASL gestures in controlled settings. These techniques, while foundational, often struggled to adapt to dynamic environments or handle complex gestures effectively.

Many studies in this category leveraged datasets containing diverse signing styles and backgrounds, demonstrating superior performance in adapting to variability in gestures and environmental conditions.

The review also discussed hybrid systems integrating multi-modal inputs, such as hand movements, body gestures, and facial expressions, which showed significant promise in enhancing recognition accuracy and robustness. By combining multiple sources of input, these systems provided a holistic representation of ASL communication, capturing subtle cues that single-modality approaches might miss.

V. RESEARCH METHODOLOGY

A. TensorFlow

TensorFlow is widely used in Machine Learning. It is a free and an open-source artificial intelligence library used for numerical data processing and to build models with the help of dataflow graphs and differential programming.

C. OpenCV

OpenCV is a huge open source library of computer vision . In the present time, it is important in real time operation and helps process images and videos to detect or identify objects, faces or even hand writing of a human being.

D. Transfer Learning

Previously well-developed model and re-used on a new set of related data. This technique is primarily used because of its advantages which are, less demanding time and data requirements.

VI. FLOW DIAGRAM OF PROPOSED WORK

Steps to be followed for hand gesture recognition are:

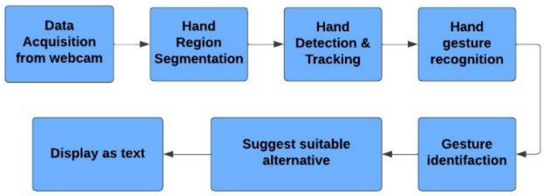


Fig. (c) Block Diagram

1)Data acquisition: Approaches to gather data are as follows-

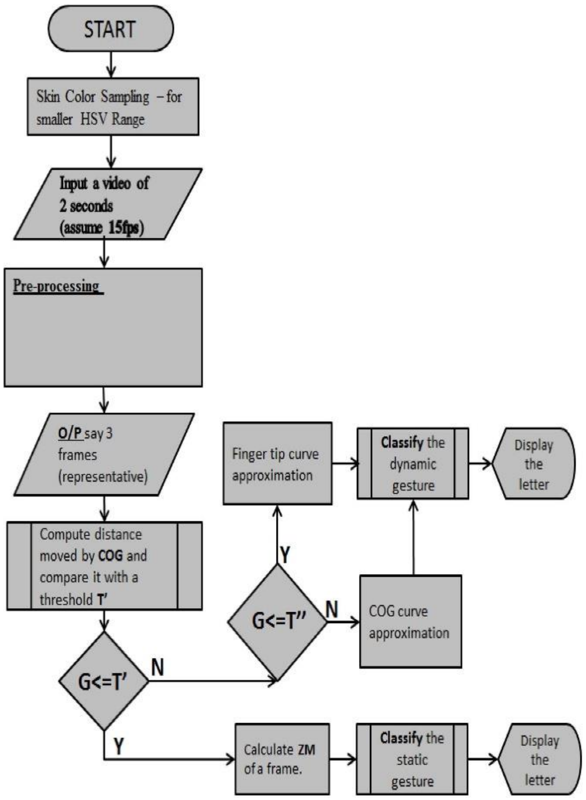
Sensory Devises – This is expensive method as it uses electromechanical devices to cut out hand configuration also various glove-based methods are used to draw out information.

Vision Approach –The major challenges faced in this method are due to different skin tones possible for human hand, movement of hand, difference in viewpoint, camera capturing speed.

when implementing such technology in professional basketball settings.

2)Data pre-processing: The background needs to be subtracted as it may contain the facial part which is of same colour as of hand. Here adaboost face detector is used to differentiate between face and hand region so that facial part could be removed. Then the gaussian blur filter is used on extracted image which is to be further trained. To get the better accuracy background of hand should be kept as single one colour as prediction and accuracy is highly dependent on lightning conditions.

3)Feature-Extraction: f the image is represented as 3D matrix the height width and depth are its dimensional properties. The value of depth of each pixel is 1 in grayscale image and 3 in RGB image.



1. **Input Gesture**: User performs an ASL sign.
2. **Capture Data**: Use cameras or sensors to record gestures.
3. **Preprocessing**: Clean and normalize the input data.
4. **Feature Extraction**: Identify key aspects of the gesture (hand shape, movement, position).
5. **Recognition Algorithm**: Match gestures to a database using machine learning.
6. **Output Translation**: Translate the gesture into text or speech.
7. **Feedback**: Allow user to confirm or correct the result.

VII. ALGORITHM USED

Convolution Neural Network Neurons have depth, width and height as 3 dimensions in comparison to standard computational neural system. The neurons are linked to window size component of the layer and each neuron is linked to such components of the input feature layer. In the fully connected region, each neuron is connected to the other. The result layer produces the single vector of class scores as the full image is reduced to its dimensions or number of classes. For object classification CNN model is considered as to be adroit. Even after applying to millions of images overfitting issue is not found in critical state, but it difficult to apply CNN model in case of high resolution images which turns out to be its drawback.

i) Convolution Layer – In convolution layer a small part of input matrix is taken which is also known as window size (generally 5*5).

values and selected region of layer is done which further produces activation matrix. Window is slided by the stride value (typically one). Thus, the network learns the filter so that whenever a particular orientation of edge/object is seen it could be activated.

ii) Pooling layer – The function of pooling layer is to minimize the activation matrix size which ultimately decreases learning parameters. Max pooling and average pooling are the types of pooling-layers. Convolution layer produces feature map and the function of pooling layer is to take out abstract of features. Feature maps generated by convolution al layer is location specific means it try to associate the particular feature with specific portion in the input image.

It further reduces the performance as it focused on gritty details .Focusing on features of higher level can solve.

The problem for that using stride with higher value could be one solution Another approach is pooling which is focused on higher level details .Max pooling and average pooling are the types of pooling-layers. Convolution layer produces feature map and the function of pooling layer is to take out abstract of features.

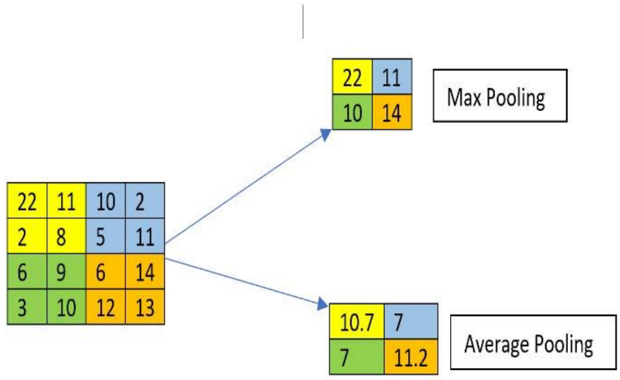


Fig. (e) Pooling

Max Pooling- n max pooling maximum of window size is picked. If window size is of 2*2 size, then the maximum value is chosen from the corresponding 4 values. Thus the new activation matrix we got is of half size than the original one.

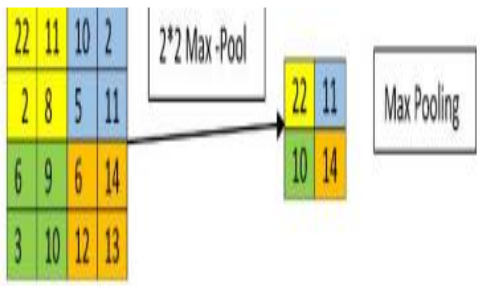


Fig. (f) Max Pooling

Average Pooling- It takes average of all values present in window.

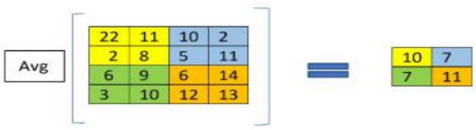


Fig. (g) Average Pooling

iii) Fully Connected layer – In convolution layer the neurons are connected only to the other neurons from the same frame but in the fully connected layer all the inputs are connected to the neuron. In order to influence every output vector with input vector it applies linear transformation to input vector. The Output from previous pooling layer is flattened and then passed to fully connected layer.

iv) Final Output layer – This is the final layer of neurons which is having the count equal to number of possible classes. Final output layer predicts the probability for the input image to be in a specific class.

v) Dropout layer: This layer abolish the effect of some neuron. It refers to set the input unit values to 0 preventing model from overfitting

IIX .ADVANTAGE OF PROPSED MODEL OVER EXISTING MODEL

Higher Accuracy Through Advanced Feature Extraction:

As compared to traditional layer built upon simple image processing methods, the suggested model utilizes sophisticated feature extraction techniques including deep learning-based convolutional neural networks (CNNs). This ensures better recognition of intricate hand gestures, even in varied lighting and backgrounds.

Real-TimePerformance:

The integration of efficient image processing techniques and optimized algorithms enables real-time ASL recognition. This is most advantageous in live communications where instant feedback is required.

RobustnesstoEnvironmentalVariations:

Current models tend to have difficulty with identifying signs under adverse conditions, such as poor lighting or occluded hands. The proposed model employs preprocessing techniques like background subtraction, edge detection,and normalization, enhancing its robustness to environmental variations.

Multilingual Support with Scalability:The proposed system is designed to accommodate a broader range of gestures and signs,making it scalable for multiple sign languages. Existing models typically focus on a limited subset of gestures, restricting their usability.

IX. RESULT AND DISCUSSION RESULT

1. Obtain gestures from the user using webcam (input).

2. Label all frames in the video to a particular word.

3. Reconstruct and display the most probable word

along with its classification scores/percentage (output)

Fig. 4. Sign Language detection model results

X. DISCUSSION:

Sign Language Recognition System has been made using Convolution Neural Networks, OpenCV andPython for the accurate detection of gesture and signs undersuitable surroundings and conditions. The model is capable oflearning all the 26 alphabets from A to Z, numbers from 0 to 9,and simple words like yes, no, hello, thank you etc. The modelcan display output in two formats, including text and audio.

The text format is accessible to the mute and the audio formatis accessible to the blind. It includes a very simple design andalgorithms to address a huge problem of the society. Here,thealphabet B has been trained into the data set and when the handgesture shows the alphabet B, the model immediately capturesthe sign and produce the desired result with 100 percentageaccuracy.

Convolutional Neural Networks have been extremely successful in image recognition and classification problems, and have been successfully implemented for human gesture recognition in recent years. In particular, there has been research in sign language recognition utilizing deep CNNs, with input-recognition that is sensitive to more than just pixels of the images. With the use of cameras that sense depth and contour, the process is made much easier via developing characteristic depth and motion profiles for each sign language gesture. The use of depth-sensing technology is quickly growing in popularity, and other tools have been incorporated into the process that have proven successful.

An effort has been placed to recognize ASL Alphabets and Numbers, which mainly depends only on hand and fingers. The process of identifying ASL Alphabets and Numbers is distributed as preprocessing the input image, computation of preprocessed image region properties, and transliteration from treated image to text

Developments such as custom-designed color gloves have been used to facilitate the recognition process and make the feature extraction step more efficient by making certain gestural units easier to identify and classify. Until recently, however, methods of automatic sign language recognition were able to make use of the depthsensing technology that is as widely available today. Previous works made use of very basic camera technology to generate datasets of simply images, with no depth or contour information available, just the pixels present. Attempts at using CNNs to handle the task of classifying images of ASL letter gestures have had some success, but using a pre-trained GoogleNet architecture.

XI. FUTURE SCOPE:

The dynamic Sign Language Recognition Model has con-siderable scopes in the future and can be incorporated even foreducation and business purposes. Some of the future scope forour Real- time Sign Language Recognition Model includes:Training the model for basic signs such as eating, drinking,walking etc. Creating a greater number of classes for numericsign languages, emotions etc. A chatbot can be incorporatedin the application for clarifying user's doubts, assisting themif they experience any issues with the application and alsofor receiving feedbacks. It can also be used in the education

sector to help children with hearing and speech impairmentsto communicate and learn through sign languages.The app can used in real-world scenarios like job interviews,meetings where the person can directly contact without needing any professional help.

We are also planning to integrateour model with applications like Google Meet, Zoom, etc.

XII. CONCLUSION

This research paper exhibits an optimal approach, to accomplish the transliteration of 24 static ASL alphabets gestures (Letter J and Z have not included as they involve hand movement. Hence it requires video frames to be processed) and 10 static ASL numbers gestures into English text.The result of our research work which proofs better recognition rate of ASL Alphabets gestures comparing with the existing traditional techniques.

This project’s primary objective is to address the societal problem of deaf and mute individuals by bridging their communication gap with hearing individuals. The basic requirement one needs to use this project is a web cam on a laptopor a mobile phone. OpenCV is being used to capture the handgesture of the signer and the data is fetched from the trained dataset model

From the results, we canconclude that this model which we made shows accuracy upto 82.6 percent under any environmental condition and evenuncontrolled lighting which addresses the major drawback inmost of the solutions that have been put forth to this problem statement. This software has a rapid processing rate and cangenerate results in real time.

REFERENCES

1. “Recognition of American sign language using image processing and machine learning” Gaikwad, Sharmila, Akanksha Shetty, Akshaya Satam, Mihir Rathod. (IEEE) (2024)
2. Varma, Teena, Ricketa Baptista, Daksha Chithirai Pandi, and Ryland Coutinho. “Sign language detection using image processing and deep learning”.International Research Journal of Engineering and Technology (IRJET) (2023).
3. Adeyanju, I. A., O. O. Bello, and M. A. Azeez. “Development of an american sign language recognition system using canny edge and histogram of oriented gradient”. Nigerian Journal of Technological Development 19, no. 3 (2022): 195-205.
4. Zhang, Chenyang, Yingli Tian, and Matt Huenerfauth. “Multi- modality American sign language recognition”. In 2022 IEEE International Conference on Image Processing (ICIP), pp. 2881-2885. IEEE, 2022.
5. Real Time Sign Language Detection- Aman Pathak, Avinash Kumar, Priyam, Priyanshu Gupta, Gunjan Chu- In-
ternational Journal for Modern Trends in Science and Tech-nology, 8(01): 32-37, 2022.
6. Sridevi, Parama, Tahmida Islam, Urmi Debnath, Noor A.Nazia, Rajat Chakraborty, and Celia Shahnaz. “Language recognition for Speech and Hearing Impaired by Image processing in matlab”. In 2021IEEE Region 10 Humanitarian Technology Conference (R10-HTC), pp. 1-4. IEEE, 2021.
7. Vij, Sugakshi, and Vivek Kumar Sehgal. ”Sign Language
Recognition Using Python and OpenCv Project.” (2021).
8. Al-Hammadi, Muneer, et al. ”Deep learning-based ap-
proach for sign language gesture recognition with efficient hand gesture representation.” IEEE Access 8 (2020).
9. Bhagat, Neel Kamal, Y. Vishnusai, and G. N. Rathna. ”Indian sign language gesture recognition using image pro-cessing and deep learning.” 2019 Digital Image Computing: Techniques and Applications (DICTA). IEEE, 2019.
10. Abraham, Abey, and V. Rohini. ”Real time conversion
of sign language to speech and prediction of gestures using Artificial Neural Network.” Procedia computer science 143 (2018).