Grenze International Journal of Engineering and Technology, June Issue

Real Time Sign Language Recognition using Knowledge Assisted Method

Vivek Srivastava[1], Jay Pratap[2], Shobhit Sinha[3] and Sunil Kumar[4]
[1]Assistant professor, ABES Engineering College Ghaziabad, Uttar Pradesh, India
Email: modest.vivek@gmail.com
[2-4]Student, ABES Engineering College Ghaziabad, Uttar Pradesh, India
Email: Jay.20b0101152@abes.ac.in, Shobhit.20b0101010@abes.ac.in, sunil.20b0101114@abes.ac.in

***Abstract*— As we are aware that verbal communication can be hampered by speech impairment, and sign language is one of the best systems for resolving this problem. The goal of our paper is to create a system or application that can recognize sign language motions in order to reduce the gap between persons with disabilities and others, such as deaf and mute people. The suggested system uses machine learning and image processing to create a real-time system. Preprocessing photos and removing hand motions from the backdrop are both done via image processing. The 24 English alphabets are then included in a dataset created using these photos. Both a specially created dataset and live hand gestures produced by people of different ages are used to evaluate the Faster(R-CNN) that we proposed in our research. The accuracy comes near to 95%.**

***Index Terms*— Deep learning, R-CNN, sign language recognition, Mute and Deaf Interaction, Recognition of Gestures, Technology for the disabled, English alphabet signs, image manipulation, computer vision.**

I. Introduction

It is impossible to overstate the importance of communication in human existence because it forms the basis of all of our interactions and expressions. Effective communication is a complex task because the ways in which we express our ideas differ greatly depending on our backgrounds, education, and societal influences. The most common form of communication for most people is spoken language, which is supplemented by body language, gestures, reading, and writing. Unfortunately, people who have trouble speaking face significant challenges because they are isolated from the general population because they only use sign language. There is a growing need for sign language recognition systems that can understand and conversely, interpret sign language into spoken or written language in order to address this communication gap. Although there are such systems, they are frequently hard to find, expensive, and difficult to use. The creation of systems for automatically recognising sign language is currently the focus of research from all over the world, with India playing a significant role in this endeavour. India, which is home to Approximately 17.7% of the global populace, is a diverse country. Surprisingly, compared to other nations, little research has been done on the recognition of real time Indian Sign Language. The lack of standardization is to blame for this delay. Indian Sign Language was first studied in 1978, but because there is no set format for it, its use is limited to short-term courses. Different deaf schools used different gestures, which made the situation worse. The standardization of Indian Sign Language and the subsequent attraction of researchers did not occur until 2003.Intricacy is a defining feature of Indian Sign Language, which

Grenze ID: 01.GIJET.10.2.253

includes signs that are both dynamic and static, single-handed and two-handed gestures, as well as regional variations for identical alphabetic characters. Recognition systems face significant difficulties due to the complexity of the absence of Indian Sign Language and a standardized dataset. In order to recognise sign language, researchers have mainly focused on two methods: the sensor-based method, which uses gloves or other tools to detect finger movements, and the vision-based method, which uses web cameras to record video or still images. The latter is preferred because it is more user-friendly, but it struggles with hand segmentation, especially in complex settings. Indian Sign Language can now be recognised effectively, precisely, and economically thanks to advancements in machine learning and deep learning. These models, which operate from beginning to end, get around limitations imposed by traditional methods, improving precision and effectiveness. This project introduces a method for creating a real-time recognition system for the A to Z and 0 to 9 Indian Sign Language alphabet and digits. This system recognises signs using webcam images rather than sophisticated technology like gloves or Kinect. The paper elaborates on the level of accuracy attained in the results. The development of a real-time, accurate, and effective system for recognising Indian Sign Language is necessary to close the gap in communication between people. with hearing or speech impairments and the general public.

II. LITERATURE SURVEY

Various authors have used a variety of methodologies, depending on the signs and sign language as a whole. A real-time sign recognition technique using Eigen weighted values Ecclesiastical distance for sign classification was introduced by J. Singha et al. **[1]**.P. Kishore et al.'s system **[2]**for classifying signs by identifying active contours from boundary edge maps made use of Artificial Neural Networks (ANN). Another method for real-time hand gesture recognition combined the Viola Jones algorithm with Local Binary Pattern (LBP) functions **[3],** which had the benefit of requiring less processing power for motion detection. A crucial manual processing step called segmentation frequently and accurately applied Otsu's algorithm **[4]**. In a related project **[5]**, initialization and segmentation steps were streamlined by using a method based on moving block distance parameterization, concentrating on high-precision static symbols and 33 fundamental word -units. Pattern recognition and feature extraction were major focuses of many of these methods **[6]**.There was an increasing demand for quicker answers to the picture identification problem in the setting of real-time systems. Through the use of several models, Deep Learning technologies played a crucial part in automating the image recognition process. In particular, Lately, Convolutional Neural Networks (CNNs) have advanced significantly. **[7,8]**. To categorize these motions and subsequently convert them into text, G. Jayadeep and colleagues **[9]** used networks with Long Short Term Memory (LSTM) in addition to a CNN to extract picture attributes. Bin et al.'s **[10]** presentation of the InceptionV3 model, which makes use of depth sensors to identify static indications, is another important strategy. The feature extraction and gesture segmentation procedures were streamlined using this model. In a different research, Using a mini-batch supervised learning technique with stochastic gradient descent, Vivek Bheda and colleagues **[11]** developed a system for classifying photographs of numerals (ranging from 0 to 9) and letters in American Sign Language. For these classification challenges, their method used deep convolutional neural networks. In their investigation, Gautham Jayadeep and associates introduced the "Mudra" solution **[12].** Mudra is a program for converting sign language to text that was primarily created to help people with voice disabilities in the banking industry. Data on sign languages were collected, subjected to preprocessing, feature extraction, Convolutional Neural Network (CNN) training, and sign classification into text. The use of a very simple dataset and the failure to take into account the nuances of common sign languages, such as complicated hand signs and factors linked owing to background and hand-body contact, seriously impair this technique. Likewise, Ismail Hakki Yemenoglu and associates have finished their study on a "CNN-based Sign Language Recognition System."

III. METHODOLOGY

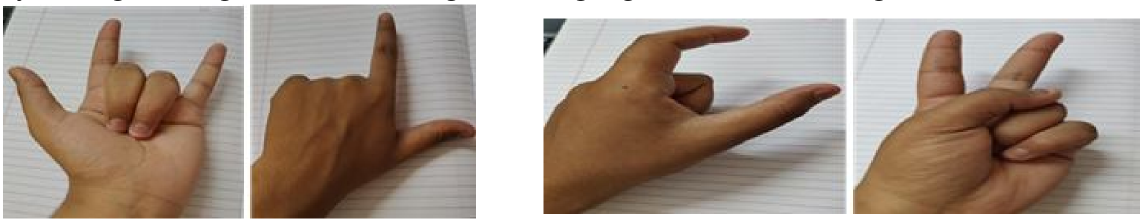
This paper outlines a method for real-time sign language detection in images. It employs a combination of a Faster R-CNN deep detector and a Residual Network-50 (ResNet-50) deep feature extractor, using 200 photos captured from various angles and resolutions with a mobile phone camera. This approach enhances the deep learning architecture for sign detection.

- *System Overview*

The study aims to identify multiple signs, and Figure 1 gives a high-level overview of the system. Below are more specifics on each stage of the suggested system.

- *Data Collection and Data annotation*

The sign-based dataset consists of 200 images total that depict various signs from the front and side. These symbols include the characters L, U, and V in addition to the words "I love you." A total of 50 image samples, including 48MP and 12MP, were taken for each sign category. A selection of various sign samples used in the suggested system are shown in Figure 1. Using the labeling tool, every image in the dataset was manually annotated by drawing bounding boxes around the signs and assigning labels to denote the signs' classes.



1. I love you word 2.Letter L 3. Letter U 4.Letter V

Fig-1.Sample images in the sign dataset

More Accurate-RCNN+ResNet50 Sign Language Identification

The objective is to identify and categorize indicators, like the words "I love you" and the letters V, L, and U. The bounding boxes around the signs must be precisely defined in order to produce accurate results. Convolutional neural networks (CNNs) are created more quickly by using the Region Proposal Network (RPN), which helps identify critical Regions of Interest (ROIs) for sign detection.

Following are the steps for a Quicker RCNN method for identifying signs in a photo:

1. An image is sent to the ConvNet to generate feature maps that will be used by the RPN.
2. The RPN creates anchor boxes with a fixed size of K of various sizes and forms within a sign image by scanning the acquired feature maps with sliding windows. It determines the bounding box adjustments required to properly fit the anchor for the sign class and forecasts the likelihood that an anchor will represent a sign.
3. A sign is added to each proposal once diverse bounding boxes of various sizes and shapes have been obtained, and Region of Interest (ROI) then collects fixed-size information maps for every anchor.
4.These compiled fixed-size attribute maps are then fed into a completely linked layer that consists of a softmax and a layer of linear regression. The detection process is finally completed after the signs are classified and the bounding boxes for the detected signs are estimated.

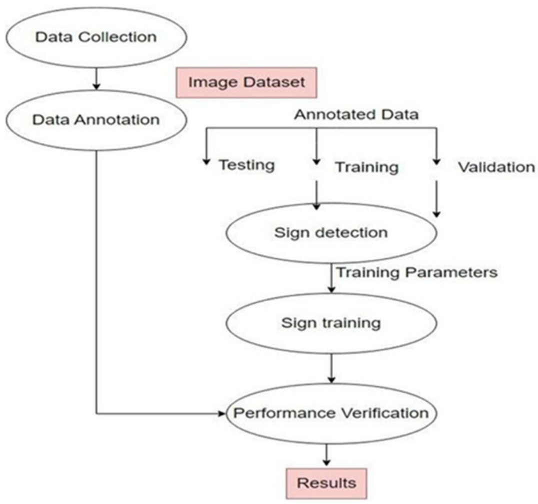


Figure 2: An overview of the suggested method for identifying sign language

IV. Experimental Outcomes

For sign detection, the recommended method makes use of Faster R-CNN and ResNet-50. What the suggested system does well is assessed using measures like IoU (Intersection over Union) and AP (Average Precision).

$Io(C, D)= |C \cap D| / |CUD|$ (1) In this case, IoU: Intersection over Union

C: Annotated ground bounding box; D: System-predicted bounding box;

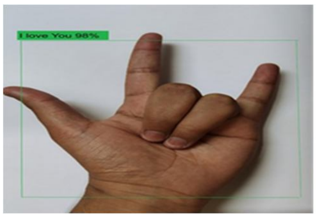
$$AP = \sum r \in \{\mathbf{0},\ \mathbf{0.1}, \ldots .\mathbf{1}\} \max p(\check{r}) \quad (2)$$

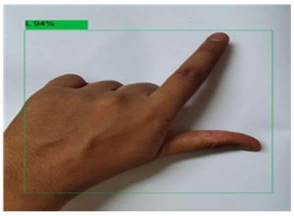
$$\check{r}: \check{r} \geq r$$

In this case, AP: Average Precision p(): Recall precision When the computed IoU value exceeds As the 0.5 predicted outcome is classed either as a false positive (FP) or a true positive (TP). The AP calculated for different signs is known as the mean average precision (mAP). Table1 displays the quantitative outcomes of the suggested system. Our suggested sign language identification system has an accuracy of 95%, and the qualitative outcomes anticipated by the proposed system are excellent.
positive (FP) or a true positive (TP). The AP calculated for different signs is known as the mean average precision (MAP). Table 1 displays the quantitative outcomes of the suggested system. Our suggested sign language identification system has an accuracy of 95%, and the qualitative outcomes anticipated by the proposed system are excellent.

TABLE I. FOR DIFFERENT SIGN CLASSES, FASTER RCNN+RESNET50 QUANTITATIVE RESULTS WERE OBTAINED.

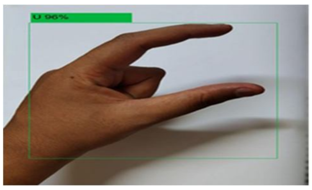
Class	Faster RCNN+ ResNet50
Letter V	92%
Letter L	94%
Letter U	96%
I Love You Word	98%
Total Mean(AP)	95%

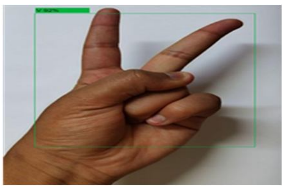




1.I Love You Word

2.Letter L





3.Letter U

4.Letter V

Figure 3: Qualitative results achieved by the system

V. CONCLUSION

Individuals who struggle with hearing or deafness frequently communicate through sign languages. Individuals who are silent or deaf use hand gestures to convey their messages, which can be difficult for people who are unfamiliar with these languages. Sign gesture recognition is largely accomplished through two methods: glove-based recognition methods and static and dynamic vision-based procedures. Although glove-based approaches are highly accurate (reaching over 98% accuracy), they may be seen as rather uncomfortable for use in the real world. We propose an approach based on vision for recognizing finger- and word-level spelling in gestures in our study. Our dataset includes 200 images of various signs, such as the words "I love you" attached to each sign, along with the letters V, L, and U. These pictures are available in two sizes: 48MP and 12MP.

REFERENCES

[1] Singha j. And das focused on the identification of live video using indian sign language in their work that was published in the international journal of computer applications in 2013.

[2] A continuous sign language recognition system using optical flow hand tracking and active contour hand shape features with artificial neural networks was presented by kishore pvv and kumar da at the ieee 6th international conference on advanced computing in 2016.

[3] An android-based indian sign language interpreter was created by swamy shanmukha, chethan mp, and gatwadi mahantesh and published in the international journal of computer applications in 2014.

[4] In 2012, feature fusion was used by jalal as, agrawal sc, and bhatnagar c to recognise indian sign language.

[5] In the journal of applied research technology, aviles-arriaga hh, sucar-succar le, and mendoza-duran ce (2011) conducted a comparative study between hidden markov models and dynamic naive bayesian classifiers for gesture recognition.

[6] A system for recognising indian sign language was created by rokade yogeshwar i and jadav prashant m and was covered in the in july 2017, the international journal of engineering and technology.

[7] The 2014 arpn journal of engineering and applied sciences published a work by sahoo a k, gouri sankar mishra, and ravulakollu kiran kumar titled "sign language recognition: current advances."

[8] Shailesh bachani, shubham dixit, rohin chadha, and prof. Avinash bagul's paper "sign language recognition utilising neural networks" was published in the international research journal of engineering and technology (irjet) in 2020.

[9] Jayadeep g, nv vishnupriya, v venugopal, s vishnu, and m geetha presented "mudra: a convolutional neural network-based indian sign language translator for banks" at the 4th international conference on intelligent computing and control systems (iciccs) in 2020.

[10] In a paper published in 2018, xy he, y. Li, and b. Xie examined "rgb-d static gesture recognition based on convolutional neural networks".

[11] The study conducted by vivek bheda and n. Dianna radpour was named "using deep convolutional networks for gesture recognition in american sign language!"" the state university of new york at buffalo's departments of computer science and linguistics undertook this study.

[12] Jayadeep gautham, vishnupriya n v, venugopal vyshnavi, vishnu s and geetha m introduced "mudra," a sign language translation tool for the banking sector, in a study published in 2020. This study was delivered at iciccs 2020, the international conference on intelligent computing and control systems.