IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 16, May 2023

Indian Sign Language Detection using CNN

Dr. Saurabh Saoji[1], Saurabh Patil[2], Prajwal Patil[3], Mahesh Rathod[4]

HOD, Department of Computer Engineering[1]

Students, Department of Computer Engineering[2,3,4]

Nutan Maharashtra Institute of Engineering and Technology, Pune, India

Savitribai Phule Pune University, Pune, India.

Corresponding Author: Saurabh Patil

Abstract*: Communication barriers often hinder the participation of the deaf community in broader society. Indian Sign Language (ISL) serves as the primary means of communication among its inhabitants. To facilitate communication between the deaf community and regular individuals, technology can be employed to convert sign languages into a comprehensible form. This paper presents a project aimed at developing a system that efficiently converts ISL into text using a deep learning technique. The proposed approach utilizes a convolutional neural network (CNN) implemented with Python-based Keras framework. The classifier model is designed to classify signs based on numerical features. In the second phase, a real-time system is employed to detect the Region of Interest (ROI) within the video frame using skin segmentation and bounding box techniques. The segmented region is then fed into the classifier model to predict the sign being performed. The system achieves an accuracy rate of 99.56% on the given topic and demonstrates 97.26% accuracy even in low light conditions. Furthermore, the classifier model exhibits improvement in performance across diverse backdrops and various picture capture angles. The proposed approach primarily focuses on utilizing an RGB camera system.*

Keywords: Deep Learning, Convolutional Neural Networks, real-time system, Computer Vision, Training, user Interaction, Indian Sign Language

I. INTRODUCTION

Both the deaf and dump communities use a variety of sign languages to communicate with persons who are physically impaired. People use a variety of languages across the world to offer communication. There are several sign languages, including American Sign Language, Chinese Sign Language, Indian Sign Language, and others. Whether motion, single-handed, or double-handed representations are present, the symbols change in each case. In certain cases, instead of using static symbols to represent letters, dynamic symbols are used for words like "hello," "Hai," etc. Through the use of a real-time technology, these communities will be able to communicate with one another. Using the Computer Vison approach, it may be modified and then translated into any language. Many studies have been done in this field in order to develop an accurate and effective method. The researchers' former method made advantage of a handmade feature, but it was restricted and only employed in certain situations.

The bulk of works use both pattern recognition and feature extraction based on HOG, SIFT, LBP, etc. However, a system with just one characteristic is frequently insufficient, which is why the hybrid approach was created to address this problem. We need to use speedier techniques to problem-solving in a real-time system, though. These days, we employ parallel implementation to speed up computer processing. The bulk of the time, the system relies on a single core to address problems. The GPU system, which has more cores than a CPU system, may be utilized to solve issues through the use of parallel computing. Using the deep learning approach, we can create a self-learning system that meets our requirements. Convolutional neural networks are among the most well-liked deep learning systems that are capable of solving any computer vision problem. We employed a region of interest-convolutional neural network to accomplish our method in real time.

DOI: 10.48175/IJARSCT-10990

II. LITERATURE REVIEW

One of the most popular trends in technology today is computer vision, which is used in many AI-based systems including robots, cars, markets, etc. The system has a greater influence on object detection and image categorization issues. This technique can be used to implant the sign language system. In the earlier systems, numerous more techniques were employed.

[1] utilized a foundation for the ISL Recognition system in literature. Color segmentation is done using a glove, and PCA (Principal Component Analysis) is employed for recognition. Every 20th frame, real-time data frames are used as input to perform recognition. This method had issues with the sign that had both overlapping and motion. PCA and the fingertip algorithm are both utilized for recognition. Recent studies have concentrated on static indicators of ISL [2] from photo or video sequences that were captured using data glove or coloured glove under controlled circumstances such a single background and specialized gear. In the system, the light and position are more significant. To work under these circumstances, the signer needs to be knowledgeable of the system.

Pre-processing Otsu's thresholding can be done in a variety of ways, including by considering skin tone, motion-based segmentation, and backdrop subtraction [3–5].Scale-invariant feature transform, Fourier descriptors, and wavelet decomposition are used in the feature extraction phase. K Nearest Neighbour (KNN), Hidden Markov Models (HMM), Multiclass Support Vector Machines (SVM)[6], Fuzzy systems, Artificial Neural Networks (ANN), and many other classifiers are used to categorize signs.

implemented an edge detection approach for hand gesture identification in another study [7].Edge detection and sorting characteristics in the database are used to retrieve the frame features. Utilize template matching to forecast the gesture using the newly built database. The smallest distance is used in this case to match templates. Both static symbols and dynamic gestures can be recognized by the system. A fuzzy membership function is used by the system to extract the spatial properties of signs utilizing a fuzzy [8] based approach. The Nearest Neighbour classifier is paired with a suitable symbolic similarity measure.

Using the Microsoft Kinect sensor device, Reheja [9] et al. devised a gesture detection system for Indian sign language. They conducted experiments using Kinect photos in RGB and Depth. According to the research, employing RGB-D photos improves the system's accuracy. The HU-Moments, which are moments that are angle, position, and shape invariant, are extracted by the model and fed to the SVM classifier as features. Indian sign language has an android app-based system designed by Pranali Loke[10]retaliates are collected by the Android system and sent to the server. The server system sends these photos to the MATLAB application, where the system is trained using a neural network and features are extracted using the Sobel operator. The system analyzes the photos using pattern recognition and classification to produce text as the result. A system to recognize American Sign Language (ASL) using the depth images captured by the Kinect sensor was created by Beena M.V. et all A total of 1000 photographs of each numerical sign were used to train the system. The approach produced a 99.46% accuracy for the depth pictures after extracting features from the block-processed images and training an artificial neural network (ANN).On the, the system has been taught for quicker execution. Convolutional Neural Network [2017-2] (CNN) with SoftMax classification is used as an extension of the work for 33 static symbols of Kinect depth images. The implementation demonstrates that the handcrafted features become insufficient for classification purposes as the number of classes rises. The CNN structure will perform better in terms of accuracy compared to other conventional methods because it can learn from the provided training data.

III. MATERIALS AND METHODS

3.1 Proposed System

Indian sign language is a sophisticated technique that uses both hands. Convolutional neural networks are applied to the image in an effective method to improve classification accuracy and for practical use. The suggested system's fundamental processes are illustrated below.

Steps

1: Enter the video frame or picture.

2. Track down the handicraft.

3. Take the feature out.

DOI: 10.48175/IJARSCT-10990

4. Sorting and forecasting.

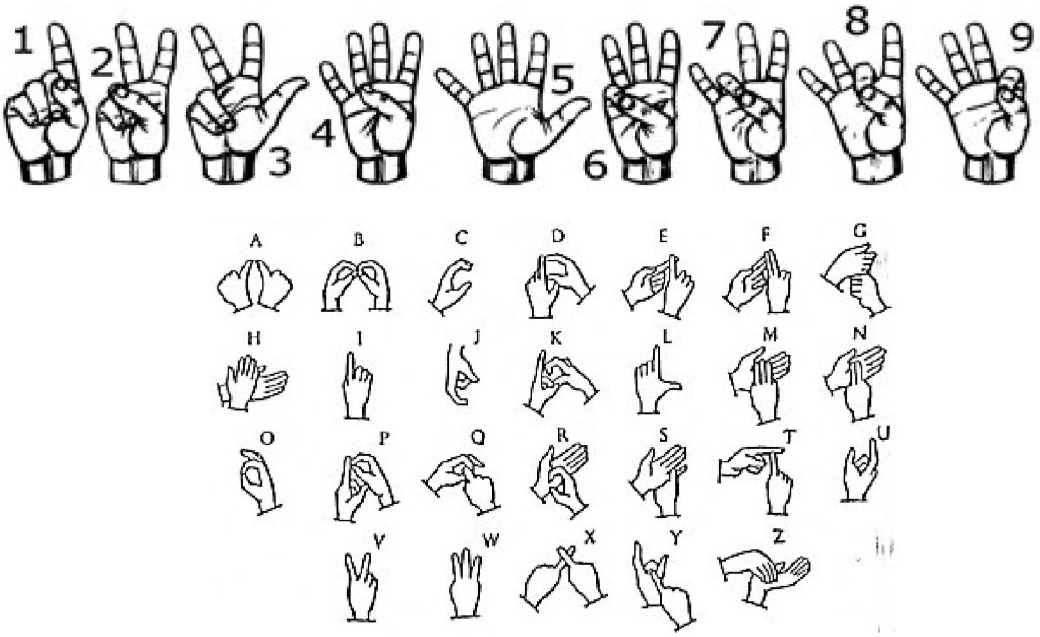


Fig 1 Indian Sign Language

The majority of object identification problems train the model using an image data set and bounding box mapping. The expense of labelling each image's bounding box is incurred. We also proposed an area of interest predictor using skin segmentation. In order to provide the classifier with a forecast, we crop the image from the segmented, limited region.

In the first phase, we input the video frame and alter the image's brightness using CLACHE (Contrast Limited Adaptive Histogram Equalisation), which uses the LAB colour scheme. The next step is to blur the original image using Gaussian blurring.

To create the skin, we do a thresholding operation using the HSV colour space. In some situations where there is a significant difference in the amount of light, we can adjust the threshold values while running. The latter steps involve identifying the largest contours in the segmented photographs and creating a rectangle box around the region that shows the output result's text classification. The bounding box is input to the convolutional neural network model to generate the text equivalent of the signs.

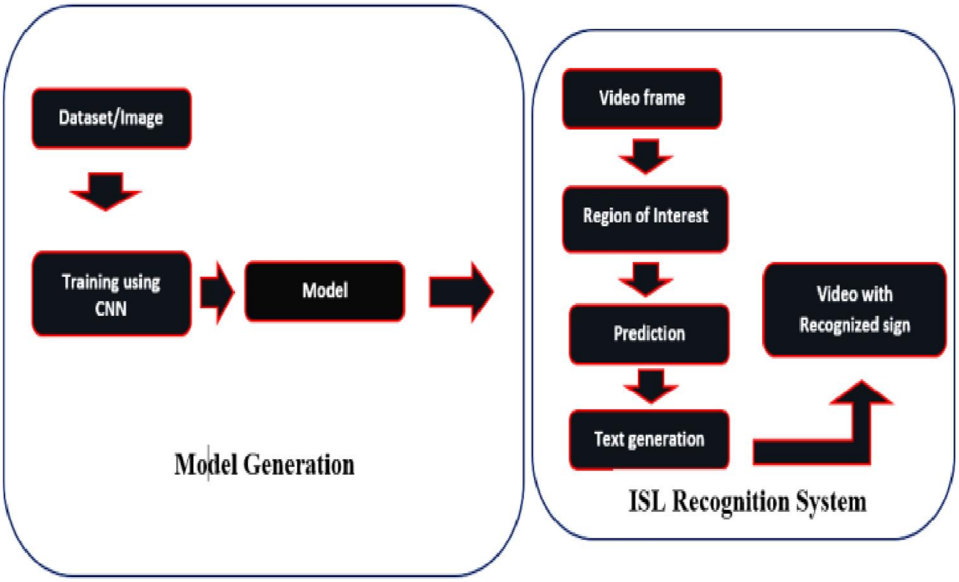


Fig: System Diagram

DOI: 10.48175/IJARSCT-10990

Algorithm :

CNN ALGORITHM:

The Convolutional Neural Network (CNN) is a widely used deep learning architecture in the field of computer vision. Computer vision, a branch of artificial intelligence, enables computers to perceive and analyze visual data or images.

Machine learning, including artificial neural networks, has shown remarkable performance in various domains. Neural networks are applied to diverse datasets, including images, audio, and text. Different neural network architectures are employed based on specific applications. For example, recurrent neural networks, particularly Long Short-Term Memory (LSTM) networks, are used for predicting word order in natural language processing tasks. Similarly, convolutional neural networks are commonly utilized for image classification tasks. In this article, we aim to provide a foundational understanding of CNNs.

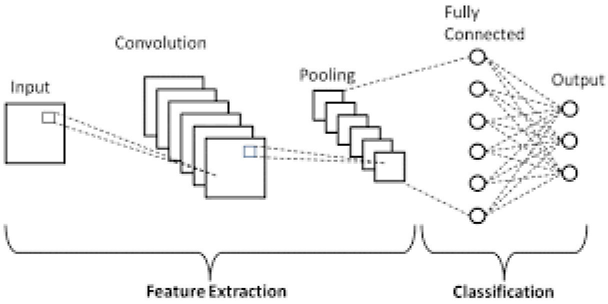


Fig CNN

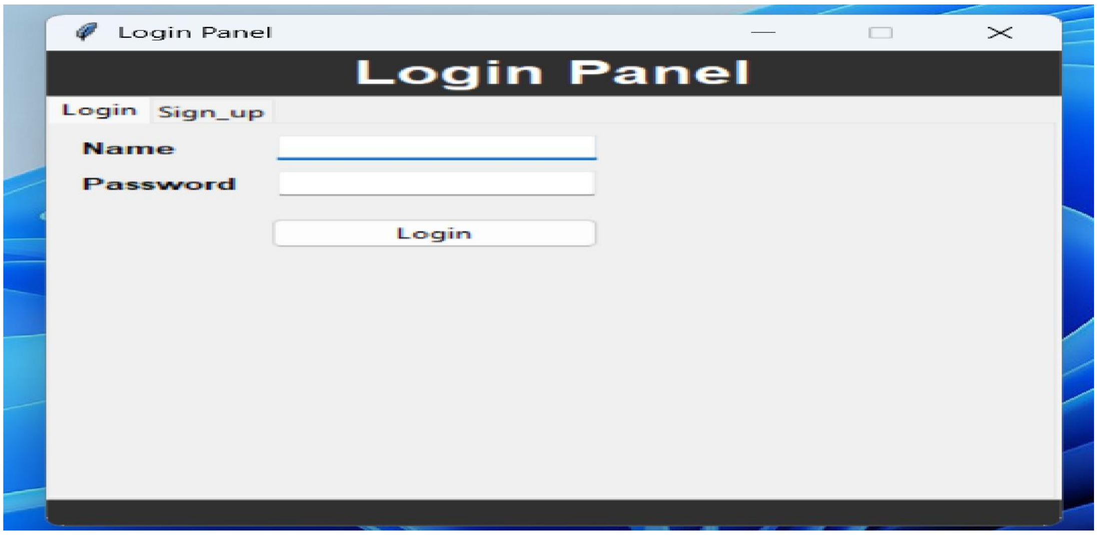
Deep Learning (ML)

Deep learning refers to a subset of machine learning that involves neural networks with three or more layers. While machine learning itself encompasses neural networks, deep learning focuses on more complex architectures. Deep learning models aim to simulate the functioning of the human brain, although they are still far from achieving a complete match. These models have the capability to "learn" from large volumes of data, allowing them to uncover intricate patterns and make accurate predictions or classifications. By leveraging deep learning techniques, significant advancements have been made in various domains, including computer vision, natural language processing, and speech recognition.

IV. RESULT

The training accuracy acquired when training the picture dataset without any augmentation was quite good (about 99%), However, the real-time performance fell short of the desired expectations. Most of the time, it not predicted accurately because hand-gestures and signs were not precisely centered and vertically aligned in real time. To address this limitation, we augmented our dataset to improve the performance of our model. Although this resulted in a decrease in training accuracy to 89%, the real-time predictions were mostly accurate. Offline testing using approximately 9000 augmented images demonstrated an accuracy of 92.7%. Moreover, as the number of parameters in the model increases, the accuracy tends to improve during both training and real-time application.

However, it is worth noting that training the model with larger image sizes requires more computational power. Due to the limited system RAM of 12 GB, we determined that the optimal image size for training was 93x63 pixels, considering that the frame size received from the Kinect sensor is 640x480 pixels. In Figure 10, we present some real-time testing scenarios using our proposed approach with various users.


Login Panel

Login Panel

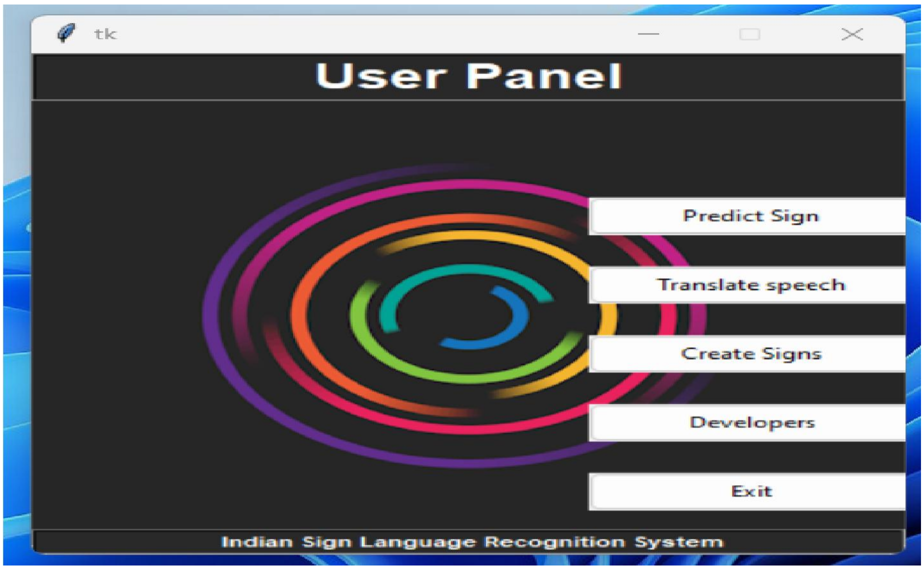
Login Sign_up

Name

Password

Login

Fig. Login Page


tk

User Panel

Indian Sign Language Recognition System

Fig. User Interface

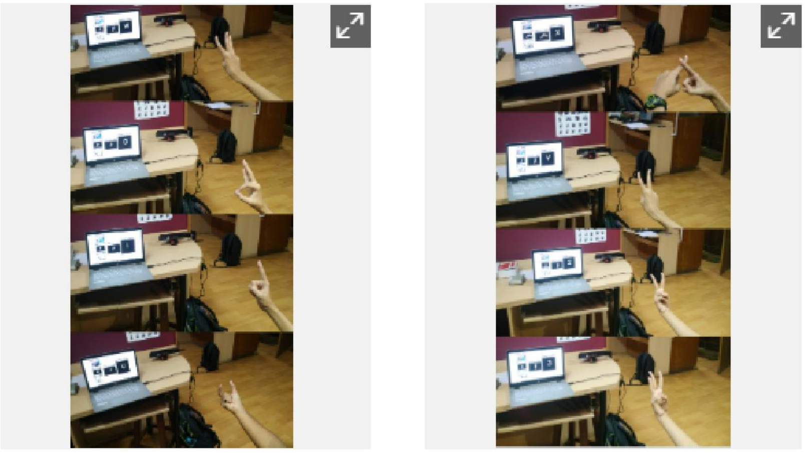


Fig. some real-time instances

Statistical tools and econometric models

HMMs (Hidden Markov Models): HMMs are a type of statistical model that may be used to represent sequential data such as sign language motions. HMMs may be taught to recognize certain sign language motions and then classified.

Support Vector Machines (SVMs): SVMs are a common classification machine learning technique. By training on a collection of examples and then categorizing new gestures based on their resemblance to the training examples, SVMs may be used to identify sign language movements.

CNNs (Convolutional Neural Networks): CNNs are deep learning models that may be used to recognize images and videos. By analyzing video data and detecting essential aspects of the motions, CNNs may be trained to recognize sign language gestures. Linear regression: A statistical model that may be used to represent the connection between two variables is linear regression. Linear regression may be used to predict the association between hand motions and specific signals in sign language recognition.

Decision trees are a widely recognized machine learning technique that can be utilized for classification tasks, including identifying sign language gestures. This approach breaks down the decision-making process into a series of simple yes or no questions. By following a path of sequential decisions based on specific features or attributes, decision trees can effectively categorize sign language motions. Each node in the tree represents a decision based on a feature, leading to subsequent nodes until a final prediction or classification is reached. Decision trees offer interpretability and can be useful in understanding the decision-making process behind sign language recognition.

V. CONCLUSION

The real-time system has been built for numeral signals from 0-9. This is the first step towards the recognition of Indian Sign Language. The 3000 static symbols of RGB images from regular camera images were used to teach the system. For testing, the system used 100 photos for each symbol. The model was developed by leveraging the power of a region-based convolutional neural network (CNN) within a deep learning system. A region-based CNN is a specialized architecture that focuses on identifying and analyzing specific regions or regions of interest (ROIs) within an image. By applying convolutional operations and pooling layers, the region-based CNN can extract meaningful features from the identified ROIs, enabling accurate classification or prediction tasks. This approach enhances the efficiency and effectiveness of the deep learning system in recognizing and interpreting sign language gestures.. For the same topic, the system achieved an accuracy of 99.56% during testing, but in low light, the accuracy dropped to 97.26%.

Add more symbols from the alphabets of the Indian sign language's static symbols in the future, including the double hand notation. The dataset must be expanded in order to address the low light issues.

REFERENCES

[1] DivyaDeora, Nikesh Bajaj, Indian Sign Language Recognition, 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication and Networking, IEEE 2012-978-1-4673-1627-9/12.

[2] Anuja V. Nair, Bindu V., A Review on Indian Sign Language Recognition, International Journal of Computer Applications (0975 – 8887) July 2013, Volume 73– No.22.

[3] Jorge Badenas, Josee Miguel Sanchiz, Filiberto Pla, Motion-based Segmentation and Region Tracking in Image Sequences, Pattern recognition 2001, 34, pp. 661-670.

[4] Ping-Sung Liao, Tse-Sheng Chen, Pau-Choo Chung, 2001, A Fast Algorithm for Multilevel Thresholding, Journal of Information Science and Engineering 17, pp. 713-727

[5] Dr. Alan M McIvor, Background subtraction techniques, Image and Vision Computing, Newz Zealand 2000 (IVCNZ00).

[6] Aseema Sultana, T. Rajapushpa, Vision Based Gesture Recognition for Alphabetical Hand gestures Using the SVM Classifier, International Journal of Computer Science and Engineering Technology, Volume 3, No. 7, 2012.

[7] Purva A. Nanivadekar, Dr. Vaishali Kulkarni, Indian Sign Language Recognition: Database Creation, Hand Tracking and Segmentation, International Conference on Circuits, Systems, Communication and Information Technology Applications, IEEE 2014,978-1-4799-2494-3/14.

[8] Nagendraswamy H S, Chethana Kumara B M, LekhaChinmayi R, Indian Sign Language Recognition: An Approach Based on Fuzzy-Symbolic Data, 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, 978-1-5090-2029-4/16.

[9] J. L. Raheja , A. Mishra, A. Chaudhary, Indian Sign Language Recognition Using SVM, Pattern Recognition and Image Analysis, 2016, Vol. 26, No. 2, pp. 434–441.

[10] PranaliLoke, JuileeParanjpe, SayliBhabal, KetanKanere, Indian Sign Language Converter System Using An Android App. ,International Conference on Electronics, Communication and Aerospace Technology, 2017 IEEE ,978-1-5090-5686-6/17.

[11] M.V. Beena and M.N. AgnisarmanNamboodiri, ASL Numerals Recognition from Depth Maps Using Artificial Neural Networks, Middle-East Journal of Scientific Research 25 (7): 1407-1413, 2017,ISSN 1990-9233.

[12] Beena M.V., Dr. M.N. AgnisarmanNamboodiri, Automatic Sign Language Finger Spelling Using Convolution Neural Network: Analysis, International Journal of Pure and Applied Mathematics, Volume 117 No. 20 2017, 9-15.