IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Review: Sign Language Recognition in ML Using CNN

[1]Advet Tomar, [2]Jay Punyani, [3]Pratik Sahare, [4]Renushree Lanjewar, [5]Harshita D. Jain

[1-4]Projecties, [5]Assistant Professor

Department Of Information Technology

Kavikulguru Institute of Technology and Science, Nagpur, India

Abstract: Deaf and dumb people all around the world use Sign Language to communicate. Yet, communication between a verbally handicapped person and a normal person has always been challenging. Sign Language Recognition is a significant advancement in assisting deaf-mute persons to communicate with others. Today, academics all around the world are concerned with promoting an affordable and accurate identification system. As a result, sign language recognition systems based on image processing and neural networks are chosen over gadget systems because they are more accurate and simpler to implement. The goal of this project is to create a user-friendly and accurate sign language recognition system that is trained using a neural network and can generate text and audio from the input gesture.

Keywords **– Hand gesture, Sign language, Communication ,OpenCV, ANN, CNN.**

I. INTRODUCTION

Deaf-dumb persons rely heavily on sign language to communicate. Each motion in sign language has a distinct meaning. As a result, complex meanings can be explained by combining numerous basic elements. Sign language is a gesture-based language used by deaf and dumb individuals to communicate. It is essentially a nonverbal language that is typically utilized by deaf and dumb individuals to communicate more successfully with one another or with normal people. Sign language has its own set of norms and syntax for efficiently communicating. There are two primary ways to sign language recognition: image-based and sensor-based. Sign language is the principal mode of communication for deaf people all over the world. A visual language that uses a different system from their everyday spoken language. Hand, facial, and physical gestures are used to communicate. Sign language is not a universal language, and multiple sign languages, like the many spoken languages spoken across the world, are utilized in various regions. Some countries, including Belgium, the United Kingdom, the United States, and India, may have several. Sign language is used. Hundreds of sign languages are used all over the world. Japanese Sign Language (JSL), British Sign Language (BSL), and Sign Language are a few examples. Turkish Sign Language is a tongue. Sign language is visual language and consists of 3 major components.

Fingerspelling	**Word level sign vocabulary**	**Non-manual features**
Used to spell words letter by letter.	Used for most of the communication.	Facial expressions and tongue, mouth, and body position.

I.1.AMERICAN SIGN LANGUAGE (ASL):

American Sign Language (ASL) is a nonverbal communication method based on the English language. It can be indicated by motions of the hands and face. It is the main language of many North Americans who are deaf or have hearing impairments. There is no universal sign language. Various nations or areas use different sign languages. For example, because British Sign Language (BSL) is a separate language from ASL, someone who knows ASL may not comprehend BSL. ASL is the fourth most regularly used language in the United States.

ASL is a language completely segregated and different from English. ASL contains all the significant features of language, with its own rules for pronunciation, word formation, and word order. While every language has ways of indicating different functions, such as asking question instead of making a statement.

I.2. CONVOLUTIONAL NEURAL NETWORK

A convolutional neural network (CNN) is a form of artificial neural network that analyses data using neural network learning rules and supervised learning. CNN is used in image processing, natural language processing, and other cognitive activities. A convolutional neural network, like other types of artificial neural networks, has an input layer, an output layer, and several hidden layers. Several of these layers are convolutional, meaning they use a mathematical model to pass findings on to subsequent layers. This mimics some of the activities of the human visual brain. CNN is a basic example of a deep learning system.

A. Input Layers: This is the layer where our model receives input. The entire number of characteristics in our data is equal to the number of neurons in this layer.

B. Hidden Layer: The input from the Input Layer is sent into the Hidden Layer. Depending on our model and data quantity, there might be numerous hidden layers. The number of neurons in each hidden layer might vary, although it is usually more than the number of characteristics. The output of each layer is produced by matrix multiplication of the preceding layer's output with the learnable weights of that layer, followed by addition of learnable biases and activation function, which makes the network nonlinear.

C. Output Layer: The hidden keyframe output is then sent into a logistic function, such as sigmoidal or max - pooling, which turns the output of each class into a probability score for each class.

II. LITERATURE SURVEY

A. Sign Language Detection and Recognition

This paper teaches us about sign language for the deaf and hard of hearing. The communication hand gesture is one technique of nonverbal communication. These tasks will help you distinguish between persons who are usually abled and those who are exceptionally abled (deaf or dumb). Because hand gestures are important in communication, the intended objective of these projects is to create a user-friendly method of communication by using the CNN (convolutional neural network) algorithm. Design: Database design, E-R Diagram, Database schema, Data flow Diagram.

These projects' keywords include CNN, Sign Language, Gesture Recognition, Open CV, ROI, Pooling, and Histogram. In this project, the input image is first recognised, and then skin colour segmentation is performed. This is the process of transforming a coloured image to a black-and-white image utilising the ROI (region of interest). ROI ignores the rest of the image and concentrates just on the skin or hand motion. The input picture is then compared to the dataset, which is provided to the model to train the system, and the input image is categorised using CNN.

TensorFlow, Keras, and OpenCV are used in this project, which is built on the Python platform. The following technological modules are employed in these projects: histogram creation, black and white conversion, gesture recognition, and letter-to-word conversion. ROI, CNN, convolution layer, RELU layer, pooling layer, and fully connected layer are the algorithms employed in these projects. The goal of these efforts is to recognise hand gestures in American sign language in real time. They trained 44 dataset samples, each with 2400 photos of a gesture, in these studies.

B. Real Time Sign Language Recognition and Speech Generation

The goal of this study is to create a user-friendly and accurate sign language recognition system trained using neural networks, resulting in the generation of text and voice from the input gesture and vice versa.

The module will utilise a computer vision-based approach to feed data from a webcam or an integrated laptop camera. The module will work on computer vision to detect hand gestures using a convolutional neural network in Python.

The module will receive input in two ways: manually collecting photos of hand positions and automatically recording video and converting it into desired frames. American sign language alphabets are a major dataset used in this curriculum.

American sign language alphabets are a major dataset used in this curriculum. The gesture datasets are pre-processed using Python libraries and packages such as OpenCV and skimage, and then trained using the CNN VGG-16 model. Speech is generated from the recognised input. As a person who does not understand sign language, this will only enable one-way communication. This module also provides text-to-sign language conversion to enable two-way communication.

The images of various alphabets of American Sign Language were collected using different webcam from different laptops. At first, a data collection program was created using OpenCV and Pillow library packages in Python The dataset is created by placing respective images for various alphabets inside folder named after that alphabet the folder name acts as the labels for training the dataset After this, the datasets for various categories were converted into NumPy array data of shape (50, 50, 3) for training the dataset using CNN model.

The ASL dataset utilised in this project's Google Collaboratory TPU training consisted of pictures of the letters A through H. The training loss and accuracy were determined to be 0.0259 and 99.65%, respectively. And 99.62% of the tests were accurate. 32,000 photos in all were utilised to train the model, while 8,000 images made up the test dataset.

C. Sign Language Recognition for The Deaf and Dumb

The use of hand gestures, facial expressions, and body movements by deaf individuals to communicate with hearing persons is known as sign language recognition. These algorithms—OpenCV, Artificial Neural Network, and Convolution Neural Network—are employed in the recognition of sign language. Machine learning and computer vision researchers are working on hand gesture detection for human-computer interaction.

Devices like vision-based systems may be controlled via hand gestures and movements for easier human-computer interaction. It is not a standard language, but it may be used to construct a variety of applications.

The goals are as follows: The prototype of sign language recognition is a real-time vision-based system that identifies American Sign Language by providing alphabets. Under adequate sun light circumstances, a camera is employed with specific limits for hand motion detection. The researchers' glove-based method is employed for 14 letter recognition, but each new user must readjust their fingers. There are two types of biometrics: physiological, which is linked with bodily form, and behavioural, which is associated with a person's behavioral traits.

Two methodologies are used by deaf and mute people i.e. Wearable communication devoce and online learning system which is glove bsed for communication. The proposed ISLR system has two modules are feature extraction and classification to joint use of Discrete Wavelet Transform. In other paper authors presented a scheme using database driven hand gesture recognition based upon skin color model approach and along with effective template and applications. In this authors presented the static hand gestures recognition system using digital image processing features by using vector SIFT algorithm having scaling, rotation and addition of noise.

Normal individuals may covert hand movements on correct text messages by utilizing Indian Sign Language (ISL).

The primary goal is to create dynamic gesture-to-text apps for the Android platform and smart phone applications. For false detection, the RGD-to-GRAY segmentation approach employs high-accuracy ISL and ASL. MATLAB is used to create the system model. For recognition, neural networks, Support Vector Machines, Hidden Markov Models, and Scale Invariant Feature Transforms are utilized. This Convolutional Neural Network is utilized in this example.

The system is founded on a vision. To reduce unwanted noise, input pictures are transformed to grayscale and subjected to Gaussian blur. The dataset problem is one of the difficulties encountered. Working with raw photos, specifically square ones, as CNN prefers to work with square images. The second difficulty is that it is difficult to choose a filter that can be applied to pictures to obtain suitable characteristics for the CNN model.

III. RESEARCH METHODOLOGY

III.1. ACQUISITION OF DATA (CAMERA INTERFACING)

This is the first and most important stage in the entire sign recognition process. Camera interfacing is required in order to capture photos using a Webcam. Several laptops now include an integrated camera system, which aids in the capture of photos for subsequent processing. The embedded camera can identify hand movements and position by capturing gestures. Collecting 30 frames per second is adequate for image processing; additional input pictures may result in longer computing time, making the system sluggish and insecure.

III.2. IMAGE PROCESSING

Image pre-processing entails eliminating undesired noise, altering the image's brightness and contrast, and cropping the image as needed [li]. Segmentation. In this process contains image enhancement, segmentation, and color filtering process.

III.3. IMAGE ENCHANCEMENT AND SEGMENTATION

Because webcam photos are RGB images, yet RGB images are particularly sensitive to light conditions, RGB information is converted to YCbCr. Where Y is the luminance information of the picture, and Cb and Cr are the chromo components that offer the color information of the image's red difference and blue difference. Because the luminance component may cause issues, only the chrominance components are processed further. Following that, the YCbCr picture was transformed to a binary image.

III.4. COLOUR FILTERING AND SKIN SEGMENTATION

As a collection of frames is captured in real time by a web camera. Because it is connected to human color perception, it is necessary to transform RGB image frames into HSV images. Color spaces are divided into three components: hue (H), saturation (S), and value (V). Image segmentation is commonly used to find hand objects and picture borders; for this purpose, HSV features allow users to specify the boundary of skin color in terms of hue and saturation value. Because V value provides brightness information, it is simple to distinguish between skin color and non-skin color information in photographs. In this method, the value of HSV is adjusted between 0 and 255 to extract and get an exact border of the object.

III.5. THRESHOLDING

The most basic approach of picture segmentation is thresholding. The thresholding procedure may be used to produce binary pictures from grayscale photographs. With thresholding, each pixel in an image is replaced with a black pixel if the intensity is less than a certain value and a white pixel if the intensity is greater than a certain value. A primary attribute that allows pixels in a picture to share their intensity. As a result, while thresholding photos, they are divided into light and dark zones.

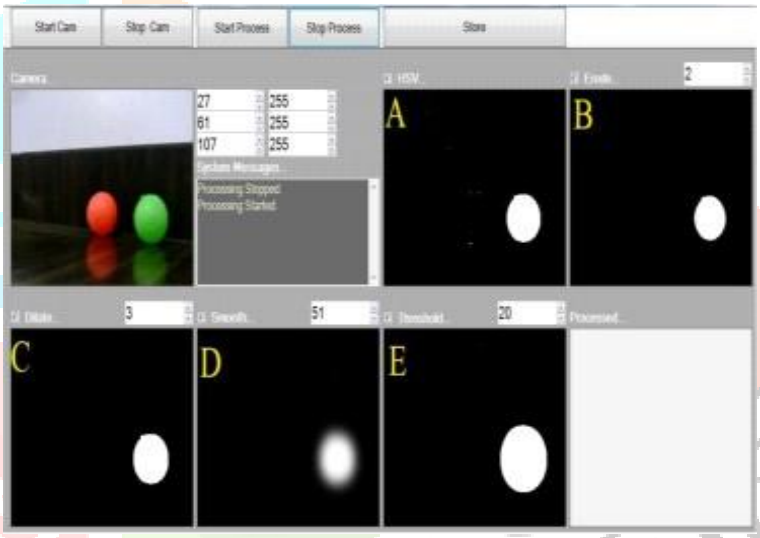

Figure 1: Image Preprocessing: Color Filtering A)HSV imageB)Erosion C)Dilation D)Smoothing E)Thresholding

III.6. CONTOUR DETECTION

The convexity hull technique is used in contour detection to construct a contour around the palm and to detect finger locations. In the convexity hull algorithm, the adaptive boosting technique is used for hand detection, and the haar classifier algorithm is used to train the classifier. The first step in the convexity hull algorithm is to segment the picture where the hand is positioned. Some feature must be assumed for this. The form of the hand is presumed here; however it may alter depending on how the hand moves. As a result, because skin color of hand is invariant to scale and movement of hand, it is considered. Separating hand pixels from non-hand pixels is the next stage in a tracking system.

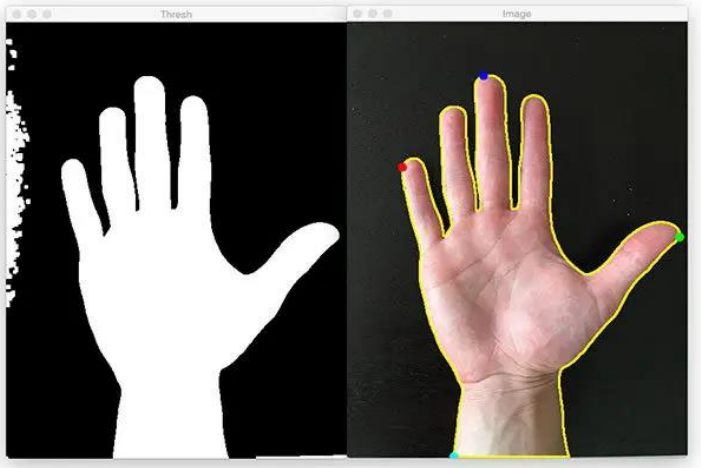

Figure2: Contour Extraction

Because sign language evolved differently than spoken language, its grammar is also distinct from spoken language. The structure of a phrase in spoken language is one-dimensional; one word follows another, however in sign language, a simultaneous structure occurs with a parallel temporal and spatial arrangement. Based on these qualities, the syntax of a sign language sentence is less rigorous than that of a spoken language phrase. A sign language phrase incorporates or refers to time, place, person, and base. A letter symbolises a sound in spoken languages. Nothing analogous exists for the deaf. As a result, those who are deaf by birth or who become deaf early in life have a very restricted vocabulary of spoken language and have considerable difficulty reading and writing.

III.7. NEURAL NETWORK DESIGN

A neural network is essentially represented by the structure seen in figure2, in which a set of components interact to create an output vector from an input vector characterized by the variable x. The neural network's collection of synaptic weight values stores the training information, and the output neuron is constrained to a specified range of activation function values.

Output neuron can be described by

$$yk=\phi(\sum wki.xi+wo), \quad (1)$$

or

$$yk= \phi(vk) \quad (2)$$

where I signify units in the input layer and k denotes hidden units; w ki denotes the input to hidden layer weights at the hidden unit k. An added that computes the weighted sum of inputs based on the weights of the connections. A node's output amplitude is defined by an activation function given an input or collection of inputs (vk), and wo is a threshold value.

In the design, a multilayer neural network with a backpropagation algorithm was utilised. The network's structure is made up of three layers: the input layer, the hidden layer, and the output layer; the fundamental components may be seen in figure 3, which uses a simplified visual notation. A hyperbolic tangent activation function with 5 neurons was used for the input and hidden layer neurons, while a linear activation function was used for the output layer neurons.

III.8. PRE-TRAINING CNN MODEL

The Transfer learning idea is employed, in which the model is first pre-trained on a dataset and then differs from the original. In this manner, the model's knowledge may be transmitted to other neural networks. The model's knowledge, which takes the form of "weights," may be stored and transferred into another model. Pre-trained models are utilized for feature extraction by layering fully connected layers on top of them. After loading the stored weights, the model is trained using the original dataset.

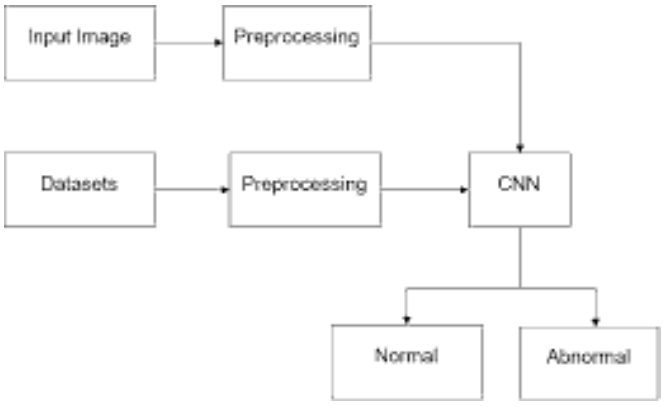

Figure 3: CNN pre-training model

IV. ALGORITHM USED

IV.1. CONVOLUTIONAL NUERAL NETWORK:

Convolutional Neural Network is a deep learning approach inspired by the visual cortex, which is the primary building block of human vision. According to the research, the human brain executes large-scale convolutions to analyse the visual data acquired by the eyes; based on this fact, CNNs are built and shown to surpass all notable classification algorithms. Convolution (wT X) and pooling (max()) are two important operations done in CNN, and these blocks are linked in a highly complicated method to imitate the human brain. The neural network is built in layers, and increasing the number of layers increases network complexity while improving system accuracy.

The CNN architecture is composed of three operating elements that are linked together to form a complicated design. The functional blocks of Convolutional Neural Network:

1. Convolutional Layer
2. Max Pooling layer
3. Fully Connected layer.

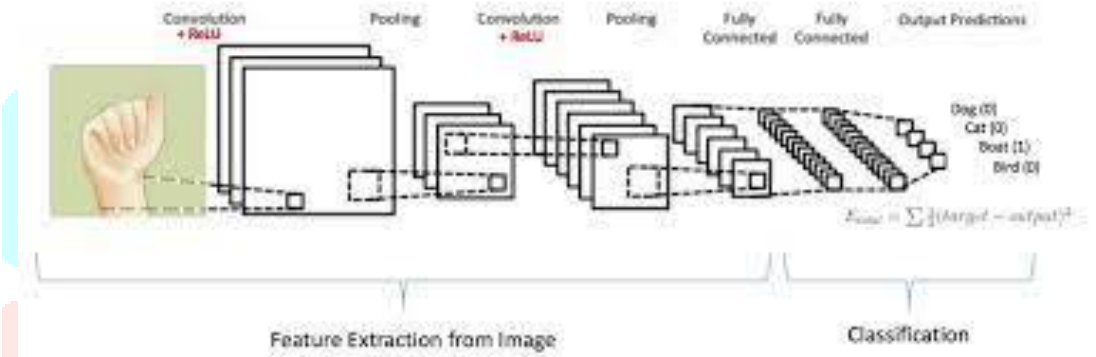

Figure 4: Convolutional Neural Network

V. CONCLUSION

The aim of the project was to recognize the Sign Language hand gestures in real time. Hand gestures are a strong way for people to interact with several applications in the field of human-computer interaction. When compared to traditional technologies, vision-based hand motion detection algorithms offer several demonstrated benefits. Unfortunately, hand gesture identification is a challenging topic, and the current study is merely a minor step towards reaching the desired results in the field of sign language gesture recognition. Videos are challenging to classify since they contain both temporal and spatial elements. To identify the spatial and temporal variables, we used two distinct models. CNN was used to identify spatial features, and RNN was used to classify temporal features. We achieved a precision of 95.217%.

REFERENCES

1. Ashok K Sahoo, Gouri Sankar Mishra and Kiran Kumar Ravulakollu 2017 “SIGN LANGUAGE RECOGNITION: STATE OF THE ART” *Asian research publishing network(ARPN),*vol. 9,no 2.
2. Mahesh Kumar N B (2018) “Conversion of Sign Language into Text” *International Journal of Applied Engineering Research ISSN 0973-4562,* Volume 13, Number 9 (2018) pp. 7154-7161.
3. Gokul Kumar K, I Imran Mohammed, Soni Jatin Mahendra 2019 “Sign Language Detection and Recognition” ***Journal of Emerging Technologies and Innovative Research (JETIR),* volume 7, issue 5.**
4. Shruti Chavan, Xinrui Yu and Jafar Saniie 2019 “Convolution Neural Network Hand Gesture Recogntion for American Sign Language” *institute of electrical and electronics engineers*.
5. Amita Thakur, Sarmila Upreti, Shirish Shrestha, Subarna Shakya 2020 “Real Time Sign Language Recognition and Speech Generation” *journal of innovative image processing (JIIP),* volume 2 pp 66-76.