Sign Gesture Recognition using Deep Learning

S Harsh [1], Viraj Bhat [2],Dr. Anitha H M [3],Dr.Shubha Rao V[4]
Dept. of Information Science & Engineering.
B.M.S College of Engineering
Bangalore, India
harsh.is20@bmsce.ac.in, virajbhat.is20@bmsce.ac.in,anithahm.ise@bmsce.ac.in

ABSTRACT

Sign language is a mechanism used by deaf and hard of hearing for their day to day communication. However, due to the limited number of sign language interpreters, communication can be a pivotal problem for those who are not acclimatized to sign language. This initiative intends to close this communication gap encountered by speech and hearing problems. This paper emphasizes mainly on creating an efficacious learning model for classification of the gestures and use of openCV for image capture. Using the approaches of deep learning such as Convolution Neural Network (CNN), this paper proposes a system to develop an application with one functionality. When fed a real-time image of an American Sign Language (ASL) hand gesture, the application processes the instantaneous image and displays the alphabet associated with it on the user's screen. The ultimate deep learning was tested and verified using various methods and was proven to be highly accurate and effective.

Keywords— Computer Vision, Gesture Recognition, Image Processing, Deep Learning, Sign Language, ASL

I. INTRODUCTION

Humans communicate with one another in many ways. This includes physical movements, facial expressions, spoken words, and other activities. However, many who have trouble hearing and speaking are only able to converse by hand gestures. People who have difficulty in hearing or speaking communicate using a common sign language that is incomprehensible to normal people. As a consequence of their impairment, learning sign language is very difficult.

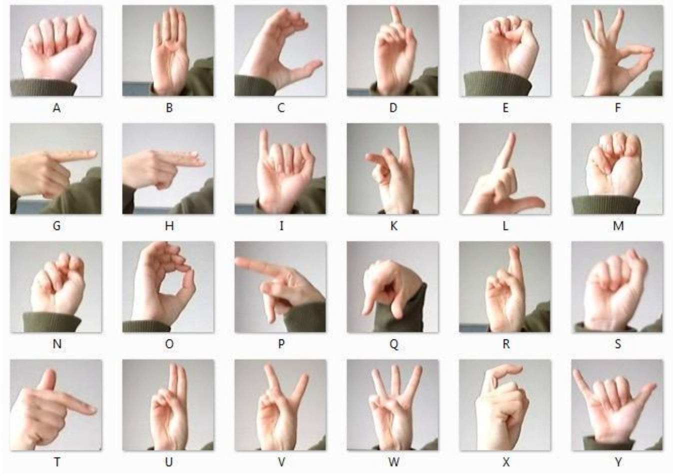


Figure 1.1 American Sign Language (ASL) [7]

The ease of sign gesture communication can be considerably improved with the implementation of a modern learning and translation tool for sign languages in machine learning. The aim of the paper is to provide following objectives:

- Fetch the video feed from the camera
- Categorize and show the equivalent English Alphabet for the ASL Alphabet.

The following are some of the challenges for this app:

1. The conditions of lighting in the area where this system is utilized must be anticipated because they have a major impact on the recognition accuracy.
2. The hand gesture needs to be at the fitting distance from the camera.
3. The camera must at least capture until the wrist and must avoid focusing on objects in the background.
4. Determining the characteristics that can be incorporated into the system to improve accuracy

This app seeks to take these intricacies into account and recognize Sign Language Symbols, with the exception of those that call for hand motion.

II. PROBLEM STATEMENT

It is very much important to interact with everyone in our modern culture, whether it's for amusement or to perform some work. Every human being needs to communicate. However, individuals with speech or hearing impairments require a different form of communication than speaking. They use sign language to converse with one another. However, learning and understanding sign language takes a lot of practice, and not everyone will comprehend what the gestures in sign language indicate. Furthermore, learning sign language takes time because there is no reliable, portable tool for doing so. People with hearing or speech impairments who are proficient in sign language need a translator who is also proficient in sign language to express their thoughts to others. This app assists people with hearing loss or speech impairments in learning and translating their sign language in order to help them overcome these issues.

III. LITERATURE SURVEY

Recognition of sign gesture is one of the cases in the research field that has been debated intermittently and is not very new. Diverse classifiers, including Bayesian networks, neural networks, and linear classifiers, have been used to solve this challenge over the past few years. Although linear models are simple to use, complex feature extraction is necessary for greater accuracy. Singha and Das' work enabled them to use Karhunen-Loeave Transforms Real-time American Sign Language Recognition with CNN to recognise photos of one-handed movements with 96% accuracy across 10 classes [1]. The input frame feed is rotated during these transformations, and a new coordinate system is established based on the data variance. Image pre-processing is used beforehand to extract the important portions of the photos. To recognize fingers pointing in a certain direction, they employ a conventional linear classifier. After background reduction and noise removal, Sharma's research uses a mix of Support Vector Machines and k-Nearest Neighbor's to identify the symbol [2]. They employ contour tracing, which simulates the curves of the hand. This method had a 61% accuracy rate. A method based on "American sign languages recognition using DL and computer vision" was proposed by Kshitij Bantupalli and Ying Xie in 2018 [3]. The authors prepared own dataset by recording films with a smartphone camera, converting each video to photos frame by frame, and then extracting 2400 images from the videos. On the dataset they produced, models such CNN employing InceptionV3 transfer learning and RNN were trained. The accuracy they attained for each sign was 91% and 55% for 150 photographs, and 92% and 58% for 50 images, respectively. A method based on the recognition of Sign language using deep learning techniques was proposed by Aditya Das et al[4]. They employed a dataset of static hand gesture portraits that were taken using an RGB camera. They trained a convolutional neural network model using the Inception v3 transfer learning algorithm on the research dataset. They attained validation accuracy of more than 90%. Using deep learning and RGB photos, Nikhil Kasukurthi et al[5]. proposed a research paper on ASL alphabet recognition in 2019.The accuracy rate achieved by authors using the deep neural network model on a squeezenet architecture was 83.29%. In 2019 there was a study published by Karlo S et al.[6] that employed deep learning to recognize static sign language. On the image dataset they used for their research, they trained CNN and the Keras method. They made use of a skin color modeling technique as the foundation for the system they created. They succeeded in recognizing static words alongside a 97.52% accuracy rate.

IV. PROPOSED SYSTEM

The android app that is being proposed only requires a camera that records video feed shown in figure 1. Each frame of this video feed is handled separately. By darkening the image and gaining the hand's white border, the outlines for the video frames are distinguishable. The hand's outline can be seen along this border. The type of symbols given in the video feed is then determined using the contours. As formerly mentioned, training the model with a dataset[11] of images containing the alphabets of Sign Language are necessary before it can be used for real time scenarios. These are given into the system after being mapped to their English alphabet equivalents.

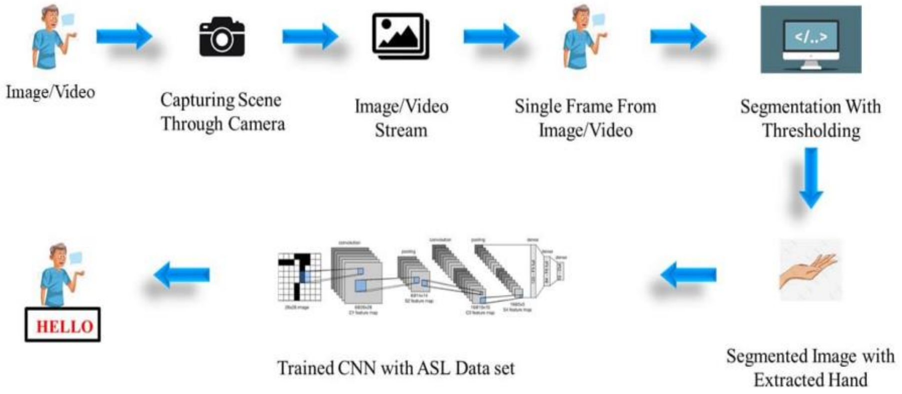

Image/Video
Capturing Scene Through Camera
Image/Video Stream
Single Frame From Image/Video
</..>
Segmentation With Thresholding
Segmented Image with Extracted Hand
Trained CNN with ASL Data set
HELLO

Figure 1. System Architecture

Taking this data into the considerations, the model is trained, and the training results are saved as a file. The model is a Convolutional Neural Network. One of the varieties of artificial neural networks (ANN) [8] is the CNN. Convolution is a mathematical procedure that is carried out by a CNN network. An input layer, an output layer, and hidden layers are the different layers that constitute a CNN. All of the layers of CNN's hidden layers conduct convolution. Three phases make up the Convolutional Neural Network's (CNN) complex design, which is connected.

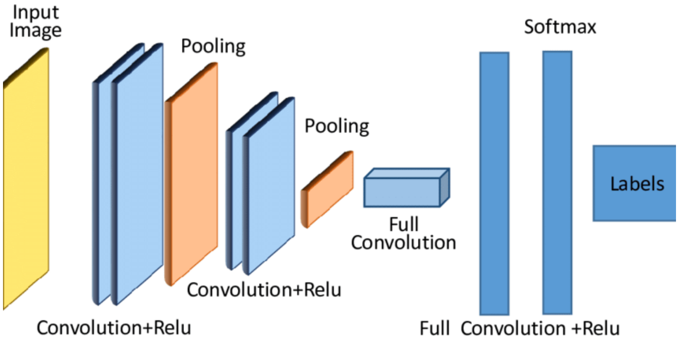
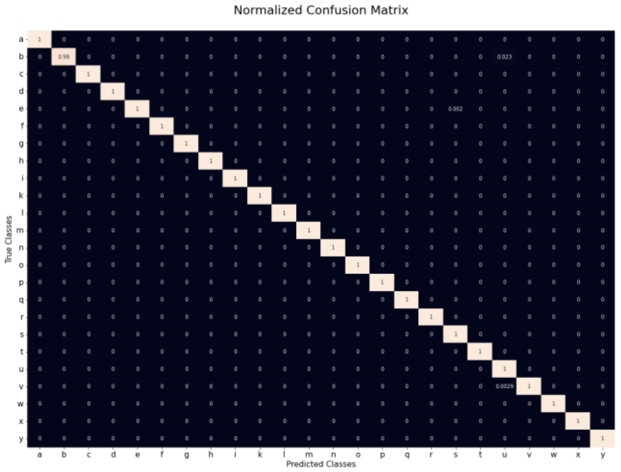


Figure 2. CNN Architecture [9]

Three convolutional layers make up the architecture shown in figure 2, and the pooling technique utilized is Max Pooling with a 2x2 pool size. The last dense layer employs the softmax activation function, while all Convolution layers employ "same" padding input sizes of (28 28 1) and "relu" activation functions. There are 11 layers altogether: First, Third, Fifth Convolutional Layers and Second, Fourth, Sixth MaxPooling Layers, Flatten Layer, Dropout Layer, Dense Layers (Relu Activation), and Output Dense Layer (SoftMax Activation) are the layers that make up this process.

V. RESULTS

The effectiveness of different models in relation to the issue at hand is evaluated using an extensive range of assessment criterias. Some evaluation measures are more suited to evaluating the performance of regression models, while others are better suited to evaluating the performance of classification models. Even though there are various evaluation metrics available, the accuracy, recall, precision, and F1 score were incorporated in this dissertation to comprehend the performance of the models.



	a	b	c	d	e	f	g	h	i	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y
a	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
b	0	0.98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.023	0	0	0	0
c	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
e	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0.002	0	0	0	0	0	0
f	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
g	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
h	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
i	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
k	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
l	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
m	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
o	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
r	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
u	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
v	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0029	1	0	0	0
w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure 3. Confusion Matrix

Figure 3 depicts the Normalized Confusion matrix where the instance of the actual class is visualized by each row and instance of the predicted class. The degree of the correctly predicted classes is represented by the values of diagonal elements.

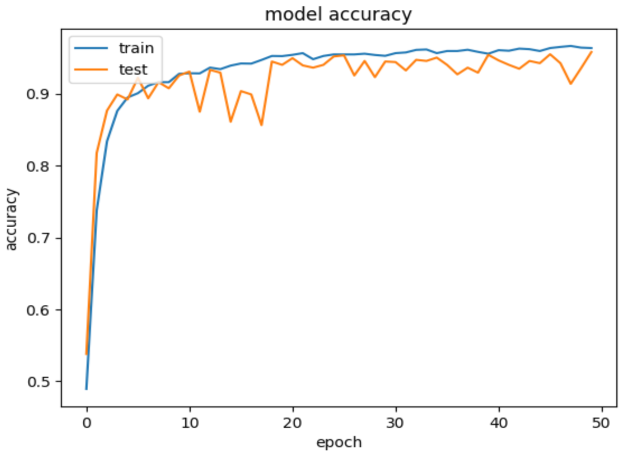


Figure 4. Model Accuracy graph

Figure 4 shows that accuracy increases rapidly for the first few epochs specifying that the model is learning fast. Later, the curve razes down indicating that with respect to further epochs the model is learning at a slow rate and has started to memorize the data.

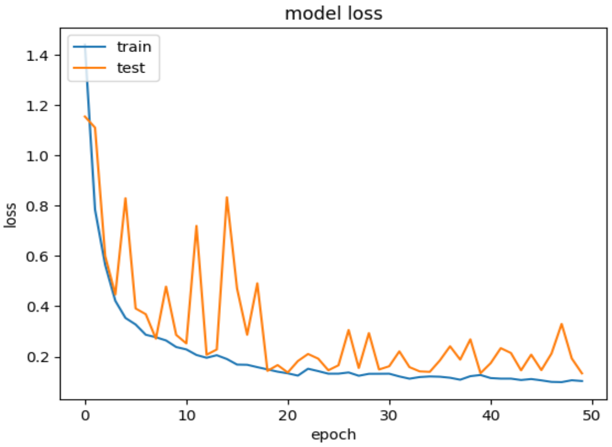


Figure 5 Model Loss graph

Figure 5 shows that, in the first few epochs, loss corresponding to the training data decreases rapidly. Then, the training data's loss does not decrease at the same rate as training data indicating the model is generalizing better to the unrevealed data.

VI. CONCLUSION AND FUTURE WORK

Convolutional neural networks and deep learning are effective techniques for a conglomeration of real-time operations, including pose recognition, multi-label image classification, natural language processing, and image classification based on labels. A white or clear background and the palm of the hand facing towards the camera are two requirements of the system. Furthermore, similar symbols could occasionally be mistaken for one another. Additionally, letters that required hand motion could not be correctly detected.

Natural language processing (NLP) [10] is a very popular field of Artificial Intelligence. With the aid of NLP, the machines are given the capacity to read, comprehend, and infer meaning from human languages such as text or speech. Consequently, applying this NLP technique will significantly enhance the application because NLP has a large potential for word and sentence recognition[12][13]. The English alphabet is employed in this project to recognize hand gestures for English letters, further opening up the possibility that phrases and sentences will also be recognized using NLP in the future.

REFERENCES

[1] Singa, J. and Das, K. " Hand Gesture Recogniton Based on Karhunen-Loeave Transform ", Mobile and Embeded 232 Technology International Conference (MECON), January 17 to 18, 2013, India. 365- 371

[2] R Sharma al. Recogniton of Single Handed Sign Language Gestures using Contour Tracing descriptor. Procedings of the World Congress on Engineering 2013 Vol. II, VVCE 2013, July 3 to 5, 2013, London, U.K

[3] Bantupalli, Kshitij, and Ying Xie. " American sign language recognition using deep learning and computer vision " In 2018 IEEE Intemational Conference on Big Data (Big Data), pp. 4896-4899. IEEE, 2018.

[4] Das, Aditya, Shantanu Gawde, Khyati Suratwala, and Dhananjay Kalbande. "Sign language recognition using deep learning on custom processed static gesture images." In 2018 International Conference on Smart City and Emerging Technology (ICSCET), pp. 1-6. IEEE, 2018.

[5] Kasukurthi, Nikhil, Brij Rokad, Shiv Bidani, and Dr Dennisan. "American Sign Language Alphabet Recognition using Deep Learning." arXiv preprint arXiv:1905.05487 (2019).

[6] Tolentino, Lean Karlo S., RO Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, and Xavier Jet O. Garcia. "Static Sign Language Recognition Using Deep Learning." International Journal of Machine Learning and Computing 9, no. 6 (2019): 821-827.

[7] "American Sign Language." NIDCD. (accessed on May 12, 2023)

[8] Visnu D.Asal and R.I Pateel, "A review on prediction of EDM parameter using artificial neural network", International Journal of Scientific Research, Vo l .2(3), 2013

[9] "Convolutional neural network." Wikipedia. (accessed on May 27, 2023)

[10] "What is NLP?" Simplilearn. (accessed on June 16, 2023)

[11] "Sign language MNIST." Kaggle. (accessed on May 13, 2023)

[12] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A. B. Gil-González, and J. M. Corchado, "Deepsign: Sign Language Detection and Recognition Using Deep Learning," Electron., vol. 11, no. 11, pp. 1–12, 2022, doi: 10.3390/electronics11111780.

[13] N. Adaloglou et al., "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," IEEE Trans. Multimed., vol. 24, pp. 1750–1762, 2022, doi: 10.1109/TMM.2021.3070438.