iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** I **Month of publication:** January 2022

DOI: https://doi.org/10.22214/ijraset.2022.39820

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

Non-Verbal (Sign Language) To Verbal Language Translator Using Convolutional Neural Network

Poonam Yerpude[1], Pratiksha Jagat[2], Reema Sahu[3], Palash Dubey[4]
[1]*Faculty CSE Department,* [2, 3, 4]*B.Tech Students, Department Of Computer Science and Engineering, Government Engineering College, Raipur, C.G. , India*

Abstract: ***Communication is very imperative for daily life. Normal people use verbal language for communication while people with disabilities use sign language for communication. Sign language is a way of communicating by using the hand gestures and parts of the body instead of speaking and listening. As not all people are familiar with sign language, there lies a language barrier. There has been much research in this field to remove this barrier. There are mainly 2 ways in which we can convert the sign language into speech or text to close the gap, i.e. , Sensor based technique,and Image processing. In this paper we will have a look at the Image processing technique, for which we will be using the Convolutional Neural Network (CNN). So, we have built a sign detector, which will recognise the sign numbers from 1 to 10. It can be easily extended to recognise other hand gestures including alphabets (A- Z) and expressions. We are creating this model based on Indian Sign Language(ISL).***
Keywords: Multi Level Perceptron (MLP), Convolutional Neural Network (CNN), Indian Sign Language(ISL), Region of interest(ROI), Artificial Neural Network(ANN), VGG 16(CNN vision architecture model), SGD(Stochastic Gradient Descent).

I. INTRODUCTION

Indian Sign Language (ISL) is used in the deaf community all over India. But ISL is not used in deaf schools to teach deaf children. Teacher training programs do not orient teachers towards teaching methods that use ISL. There is no teaching material that incorporates sign language. Parents of deaf children are not aware about sign language and its ability to remove communication barriers. ISL interpreters are an urgent requirement at institutes and places where communication between deaf and hearing people takes place but India has only less than 300 certified interpreters[1].

"Sign languages (also known as **signed languages**) are languages that use the visual-manual modality to convey meaning. Sign languages are expressed through manual articulations in combination with non-manual elements"[10].



Figure 1: (1-10) Numerals in Indian sign language

The above image fig1 of the numbers was created from a live camera feed.The main 2 ways in which we can convert the sign language into speech or text i.e. , 1) Sensor based technique, 2) Image processing[2][3]. The above techniques are also mentioned in paper [4]. Sign language recognition processes have various types, but there are quite a bit of similarities in them [3]. The three main and common steps of this process are Pre-processing, Feature extraction and classification. There are many techniques for image processing [5]. Support vector machine (SVM) is also one of the techniques that can be used for feature extraction [6]. Multilayer perceptrons are used to apply computer vision, now succeeded by CNN[7]. MLP is now considered insufficient for modern advanced computer vision tasks. As it contains the properties of a fully connected layer, in which every perceptron is connected with each other perceptron. The disadvantage of this is that it causes the total number of parameters to grow very high (number of perceptrons in layer 1 multiplied by number of perceptrons in layer 2 multiplied by numbers of perceptrons in layer 3 and so on ...). This is inefficient due to redundancy in such high dimensions, It also disregards spatial information, which is also a con.

©IJRASET: All Rights are Reserved

Convolutional neural network is a class of neural networks in deep learning. It is mainly used in processing data with a grid-like topology, e.g. images. A digital image is a binary representation of visual data. Since MLP is a fully connected network, it makes them prone to overfitting data. Thus we are using CNN for this sign language recognition model. This project has 3 steps -

1) Creating a database.
2) Training the CNN model.
3) Predicting the data.

We are using following technologies -
Python, Jupyter Notebook(IDE), Keras, Tensorflow, openCV, Numpy.
We will capture the live feed for each hand gesture from the camera using openCV and store it in train and test folders. Then we will train the CNN model with this train data. In the 3rd step we will predict the results using test data. After training our model we got an accuracy of about 89.9 %.

II. METHODOLOGY

In this project we are creating a sign detector, for ISL alphabets, and numbers (1 - 10). For this, we are using the CNN algorithm for training the model, Keras, which is an open-source software library that provides a python interface for ANN and acts as an interface for the TensorFlow library, openCV for capturing the real time feed from a camera. As discussed above the project has three steps. Thus in the first step we are creating a database for hand gestures. The openCV library is used for the following. In the 2nd step we are feeding this data to our CNN model which will give us an 1D array each for a hand gesture, each array will contain the details which specifies a particular hand gesture. We will train our model with this data and then predict the outcome for the captured hand gestures.

A. CNN

Definition - "The name "convolutional neural network" indicates that the network employs a mathematical operation called convolution[11]. Convolutional networks are a specialized type of neural networks that use convolution in place of general matrix multiplication in at least one of their layers."
A typical CNN consists of mainly three layers, that are - Convolutional, Pooling layer and Fully connected layer. There have been many endeavours in improving the CNN for large-scale image and video recognition[7]. VGG 16 is a convolution neural network (CNN) architecture [8] . It was proposed by Karen Simonyan and Andrew Zisserman of the Visual Geometry Group Lab of Oxford University in 2014. This CNN model is considered to be the best vision architecture model till date. VGG 16 architecture's main focus is on having convolution layers of 3x3 filter with a stride 1 and always uses the same padding and maxpool layer of 2x2 filter of stride 2, instead of having a large number of hyper-parameters. Throughout its whole architecture it consistently follows this arrangement of convolution and maxpool layers and in the end has 2 fully connected layers. The following layers are given below. 1. Convolutional layers 2. Pooling layers. 3. RELU(Rectified Linear Unit) 4. Fully connected layer 5. Softmax layer. The architecture is also shown in the figure 2 below:

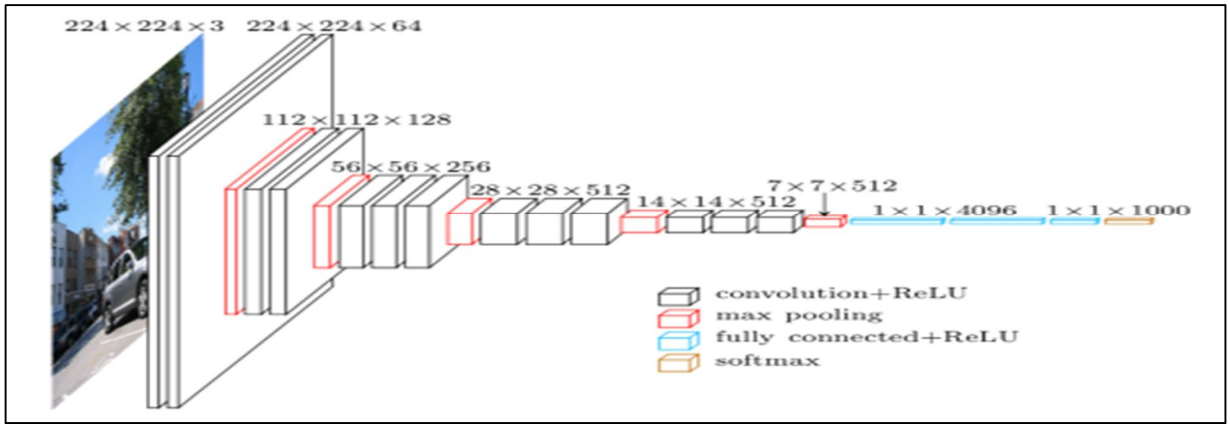


Figure 2 : VGG 16 Architecture

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

B. Implementation

We are using this model in our project. So first thing we will do is, create an object of the ImageDataGenerator for training as well as testing data and pass the address of folders to the respective objects. In our directory structure there are two folders one is train, and other is test. Both folders contain 10 folders which belong to the sign language hand gestures from 1 to 10 respectively. The ImageDataGenerator will automatically label the data inside the folders as 1,2,3, and so on. Now our data can be feeded to the neural network. The directory structure is shown below in figure 3:

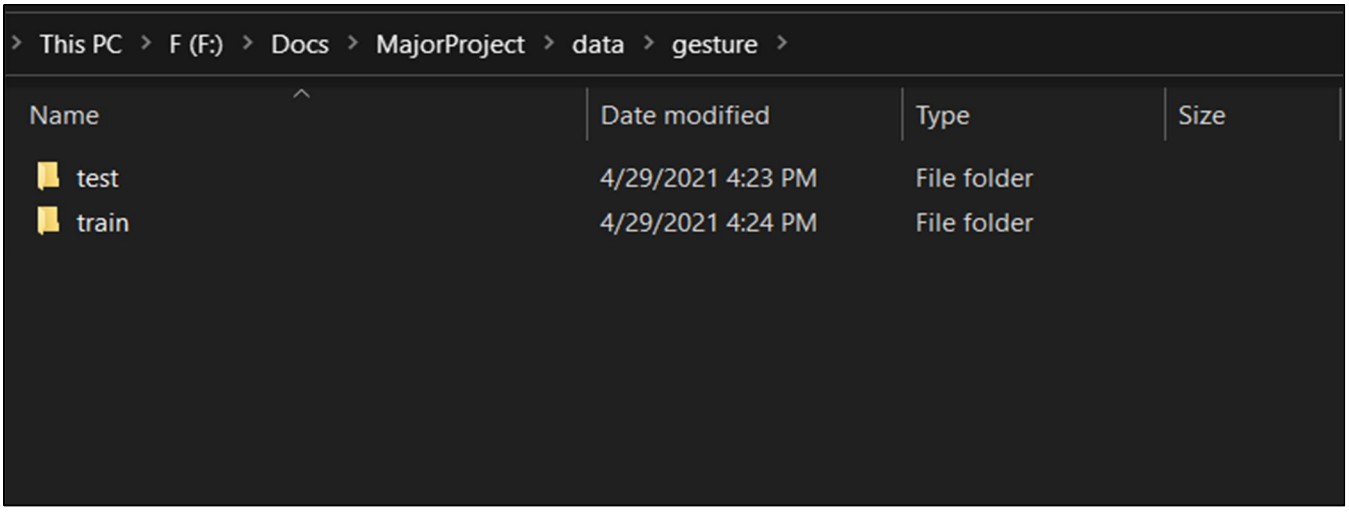


Figure 3 : Directory Structure

Now, we have added the convolution and pooling in the following order:

1) 1 convolution layer of 32 channel of 3x3 kernel and with input shape of 64x64x3
2) 1 maxpool layer of 2x2 pool size and stride 2x2
3) 1 convolution layer of 64 channels of 3x3 kernel and with the same padding.
4) 1 maxpool layer of 2x2 pool size and stride 2x2
5) 1 convolution layer of 128 channels of 3x3 kernel and with valid padding.
6) 1 maxpool layer of 2x2 pool size and stride 2x2

All the above layers have RELU(Rectified Linear Unit)Activation function. The relu can be defined as shown in figure 4 :

$$f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$$

Figure 4 : RELU Function

The pooling technique we are using is max pooling [7]. After this we have implemented the dense layer as:

a) 1 Dense layer of 64 units with relu activation
b) 1 Dense layer of 64 units with relu activation
c) 1 Dense layer of 128 units with relu activation
d) 1 Dense softmax layer of 10 units.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

At the last, we are using a Dense softmax layer of 10 units, as there are 10 classes from 1to 10 to predict. The softmax function can be defined as show by the equation in figure 5:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

σ = softmax
$\vec{z}$ = input vector
e^{z_i} = standard exponential function for input vector
K = number of classes in the multi-class classifier
e^{z_j} = standard exponential function for output vector
e^{z_j} = standard exponential function for output vector

Figure 5 : Softmax function

The model is complete after this. There are 2 different optimization algorithms that can be used – SGD (stochastic gradient descent, in this the weights are updated at every training instance) and Adam (combination of RMSProp and Adagrad). We found that our model was giving more accuracy when we used SGD as compared to Adam.

III. PREDICT THE GESTURE

Here, we create a bounding box for detecting the ROI and calculate the accumulated_avg for identifying any foreground object, as we did in creating the dataset. The figure 6, below shows the predictions made by our model:

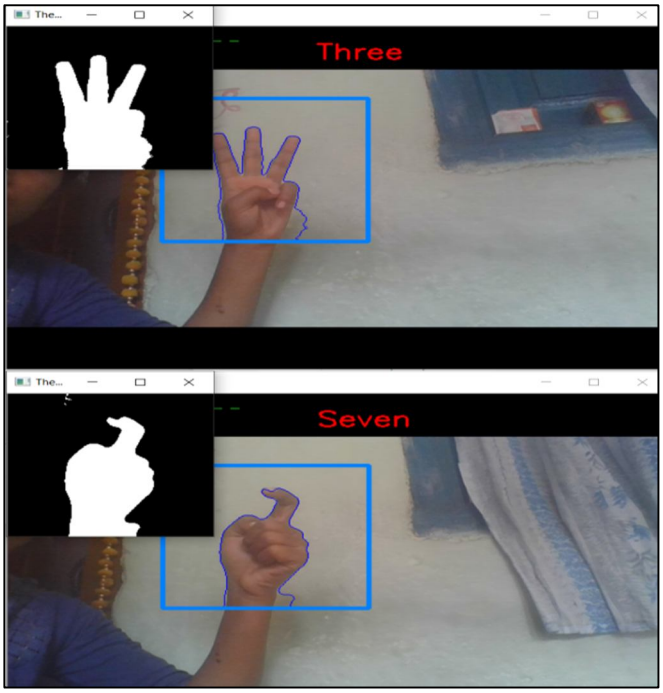


Figure 6 : Predicting the gesture

Now we find the max contour and if contour is detected that means a hand is detected, so the threshold of the ROI is treated as a test image.We load the previously saved model using keras.models.load_model and feed the threshold image of the ROI,which consists of the hand gesture as an input to the model for prediction [9].

©IJRASET: All Rights are Reserved

IV. CONCLUSION

The main goal of our project is to reduce the communication barrier between the normal and disabled(deaf and dumb) people. We have focused on translating the ISL (Indian Sign Language) numerals (1 to 10) into English alphabetical representation. This application will lead to a decrease in shortage of interpreters and trainers in ISL. We have used the VGG 16 Architecture for our model, which is providing an accuracy of 89.9%. As we are using the VGG 16 architecture, the performance of our model is better as compared to other models.

For future work we will be translating the ISL alphabets to English alphabets and the English alphabets to ISL.

V. DECLARATIONS

A. *Study Limitations*

Dataset - as we have generated the dataset using our camera apparatus and lighting conditions, the trained model may have different accuracies for different lighting conditions and camera specifications.

B. *Competing Interests*

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] History- Indian Sign Language Research and Training Centre (ISLRTC) Department of Empowerment of Persons with Disabilities, Divyangjan. http://www.islrtc.nic.in/history-0#:~:text=%20Indian%20Sign%20Language%20%28ISL%29%20is%20used%20in,is%20no%20teaching%20material%20that%20incorporates%20sign%20language

[2] Anju Varghese , Christy Paul , Dilna Titus , Vijith C, "A Survey on Sign Language to Verbal Language Converter", International Journal of Engineering Science and Computing (IJESC), September 2016, Volume 6 Issue No 9. https://ijesc.org/upload/afd13342e796097ee236bf9916722409.A%20Survey%20on%20Sign%20Language%20to%20Verbal%20Language%20Converter.pdf

[3] Daleesha M Viswanathan , Sumam Mary Idicula,"Recent Developments in Indian Sign Language Recognition: An Analysis" , International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 6 (1) , 2015, 289-293 http://ijcsit.com/docs/Volume%206/vol6issue01/ijcsit2015060165.pdf

[4] Hema B N, Sania Anjum, Umme Hani, Vanaja P, Akshatha M, "Survey on Sign Language and Gesture Recognition System", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 03 | Mar 2019. https://www.irjet.net/archives/V6/i3/IRJET-V6I3766.pdf

[5] Yellapu Madhuri, Anitha.G, Anburajan.M "VISION-BASED SIGN LANGUAGE TRANSLATION DEVICE ", Conferenc Paper · February 2013 https://www.researchgate.net/publication/261460857_Vision-based_sign_language_translation_device

[6] Prof. Radha S. Shirbhate, Mr. Vedant D. Shinde, Ms. Sanam A. Metkari, Ms. Pooja U. Borkar, Ms. Mayuri A. Khandge, "Sign language Recognition Using Machine Learning Algorithm ", International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 03 | Mar 2020 https://www.irjet.net/archives/V7/i3/IRJET-V7I3418.pdf

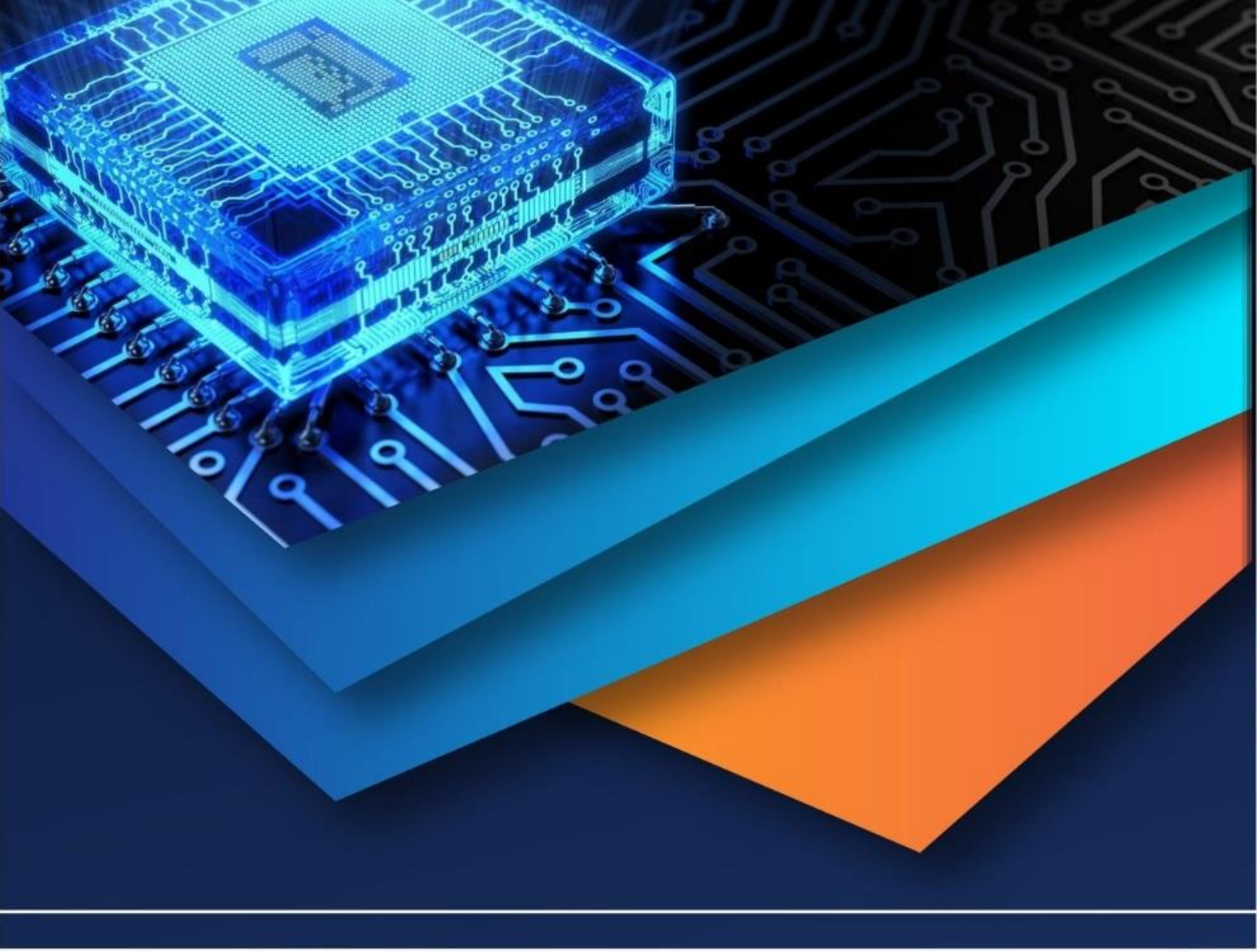
[7] Abdelaziz Botalb, M. Moinuddin, U. M. Al-Saggaf, Syed S. A. Ali, "Contrasting Convolutional Neural Network (CNN) with Multi-Layer Perceptron (MLP) for Big Data Analysis ", 2018 International Conference on Intelligent and Advanced System (ICIAS). https://www.researchgate.net/publication/329493783_Contrasting_Convolutional_Neural_Network_CNN_with_Multi-Layer_Perceptron_MLP_for_Big_Data_Analysis

[8] Karen Simonyan & Andrew Zisserman ,"VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION", Published as a conference paper at ICLR 10 Apr,2015 https://arxiv.org/pdf/1409.1556.pdf

[9] Sajanraj T D, Beena M V, "Indian Sign Language Numeral Recognition Using Region of Interest Convolutional Neural Network ", Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE. https://www.researchgate.net/profile/Sajanraj-T-D/publication/327937028_Indian_Sign_Language_Numeral_Recognition_Using_Region_of_Interest_Convolutional_Neural_Network/links/5c8355b7299bf1268d488652/Indian-Sign-Language-Numeral-Recognition-Using-Region-of-Interest-Convolutional-Neural-Network.pdf

[10] Wikipedia-"Sign-Language", https://en.wikipedia.org/wiki/Sign_language

[11] Wikipedia-"convolutional neural network", https://en.wikipedia.org/wiki/Convolutional_neural_network



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)