ISSN 2278-3091

Volume 10, No.4, July - August 2021

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse291042021.pdf

https://doi.org/10.30534/ijatcse/2021/291042021

Sign Language to Text Conversion Using Deep Learning Techniques

Gadhi Sreenivasa Reddy, Dr.Parvathi.R
Vellore Institute Of Technology, Chennai

ABSTRACT

Sign language is one of the oldest and most people used to communicate with people like Deaf and Dumb. Presently most people don’t know sign language and it’s become very tough to communicate with deaf people to clear those barriers. They created a real time live fingers spelling based application using Deep learning techniques. Gesture is the Symbol and physical emotion to the Deaf people. In this Method first hand is passed through the Filter and based on fingers it will predict the Alphabet and Display. Here I was using the VGG16 model which gives 99% accuracy for the 26 letters of the Alphabets.

Key words: Sign Language Character Recognition, Convolution Neural Network,Visual Geometry Group from Oxford(VGG-16),Residual Networks is a classic neural network(RESNET-50),Computer Vision, Deep Learning.

1.INTRODUCTION

Communication is the process which plays a vital role in our lives to share ideas and feelings. There are people who born with Deaf and Dumb and some will become disability due to some circumstances. Most of the 60-80% in the world people using English language to communicate. People used hand gestures to communicate with people with disabilities but from past years many people don’t know about Sign language(hand gestures) to communicate with them. Hand gestures are the nonverbally exchanged messages and these are understood by the vision.My aim is to create a Computer application which helps to convert sign language to American Language to help people. Here I was creating a data set from my own images and training dataset with multiple algorithms finding the best algorithm which gives best accuracy from that i was choosing an algorithm to build the Application and Deploying it into the cloud.

Various Image processing techniques and machine learning based models using SVM, Random forests-based approaches have been proposed to completely automate the cephalometric landmark identification. Recently, deep neural network-based architectures achieved a wide popularity in the field of medical imaging and obtained better classification and prediction accuracy compared to other existing technologies.

The people who are not deaf they don’t have an interest in learning any sign language. But people may have some relations with deaf people. If the computer is programmed in such a way to understand any language and help to fill those gaps. Most of the people know English more than other languages like Telugu, Tamil, Malayalam, French etc... There are people who born with Deaf and Dumb and some will become disability due to some circumstances. Most of the 60-80% people in the world use English language to communicate. People used hand gestures to communicate with people with disabilities but from past years many people don’t know about Sign language (hand gestures) to communicate with them. Hand gestures are the nonverbally exchanged messages and these are understood by the vision.

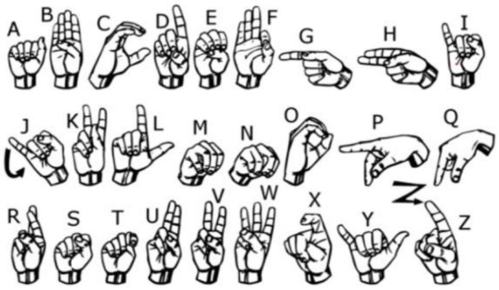


Figure 1: Sign language Alphabets

2. LITERATURE SURVEY

Over the past few decades, lots of research is going on in gesture recognition as it can be used in various application domains like smart home applications, Human Computer Interaction (HCI), gaming, medical systems, etc. Solutions proposed by different researchers are of two types: solutions based on Hardware and solutions based on Soft- ware. Solutions based on hardware include gesture recognition using gloves, wrist bands, etc. These hardware solutions contain sensors as they are necessary to track hand movements. Google has developed wristbands which are able to recognize gestures by track- ing hand movements and user is able to hear the recognized word/sentence through a mobile device as the mobile device is connected to the wristband [7].Glove based solutions are also developed in recent years. CyberGlove was unable to detect all fingers associated with ASL gestures because of the limited number of sensors. Because of the number of sensors, CyberGlove was not able to differentiate between some gestures in which wrist positions are almost similar e.g., R and U, G and H, etc. [3]. In another proposed method, data captured by gloves is sent to the neural network and is processed for classification [4]. InerTouchHand System is proposed for Human Machine Interaction (HMI) and uses distributed inertial sensors, vibro tactile simulators [8]. Glove based systems may give wrong results as time goes on depending on sensor quality. Software based solutions include gesture recognition using Support Vector Machines(SVM), Neural Networks(NN), Hidden Markov Models (HMMs), etc. Software based solutions require image processing before classifying gesture images. Amazon Alexa also is able to respond to sign language gestures[4]. But in this system, you have to capture yourself repeatedly performing each sign every time you launch the site in thebrowserandthisisaverytedioustask.Also,thesesystemsarenotaffordablebyall people. Histogram of Gradients (HOG) and Scale Invariant Feature Transform (SIFT) features are drawn out from the images of hand gestures and are fed to Support Vector Machines (SVM) for training which is then used to classify new hand gesture images [9]. They used a dataset containing images of different orientations for accurate classification. HOG along with Local Binary Pattern (LBP) features are used together to classify hand gestures and this system attained an accuracy of 92% [5].

3.IMAGE AUGMENTATIONAND RESIZE

The image dataset consists of ASL gestures from. The dataset consists of 2080(26*80) images with 80 images per category. Each category represented a different character of English Alphabet. This dataset was then augmented to create a dataset of 1781 images. Out of this dataset 75% i.e. 1500 images were used for training and remaining 25% i.e. 580 images were used for testing.

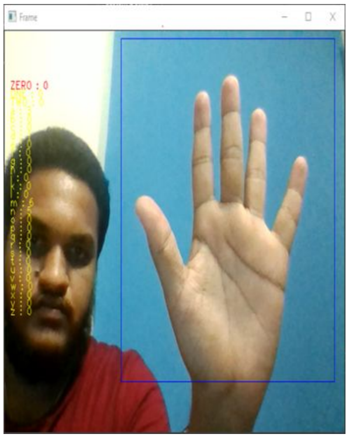


Figure 2: Taking user input

The images in the data set were of a varying size and shape. Therefore the first step was to read and resize each of the images to the similar size of 224x224 pixels. Only when all of the images in the dataset are of the same size can the images be fed into a neural network for training.

4.IMAGE PREPROCESSING

The mean value of RGB over all pixels was subtracted from each pixel value. i.e in the first pass the model will compute the mean pixel value of each channel over the entire set of pixels in a channel and in the second pass it will modify the images by subtracting the mean from each pixel value. Subtracting the mean value from the pixels centres the data. The mean is subtracted because the model involves. Multiplying weights and adding biases to the initial inputs to cause activations then back propagated with the gradients to train the model.

For Image Pre-processing I used the Direct Python function to convert the colour image to gray and remove blur which was used in this project and saving gray images in another directory.

```
frame=cv2.imread(path)
gray=cv2.cvtColor(frame,cv2.COLOR_BGR2GRAY
)
blur = cv2.GaussianBlur(gray,(5,5),2)
```

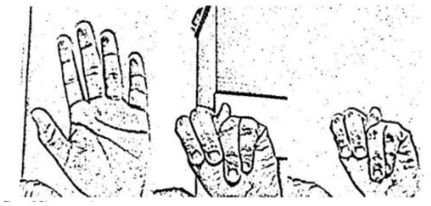


Figure 3: Converting color image to gray scale image

5.IMPLEMENTATION

Convolution Neural Network Model(CNN)

The Convolution Neural Network Model was proposed by LeCun, and has made a breakthrough in the field of image classification and target detection. Deep CNN‟s introduces a large number of hidden layers, thus reducing the dimensionality of the image and enabling the model to extract sparse image features in low dimensional space.

Now these images are reshaping into 126X126 and passing these images for model building.

The first layer is a convolution layer with 32 filters with activation function as rely, this is followed by max pooling layer with a polling size of 5,again I have used the same convolution and max pooling layer now a dropout has been introduced so that the model don’t get over fitted, after this layer I have put an convolution layer with 32 filters with activation function as rely followed by max pooling layer, same layer has been repeated again now a flatten layer has been introduced and since we have 26 classes to classify, last layer is dense layer with 26 node with activation function has SoftMax.

After building the model I have compiled the model by using Adam optimizer, loss function as categorical cross entropy and evolution matrices as accuracy.

Visual Geometry Group from Oxford(VGG-16)

VGG 16 model is a deep convolutional neural network model proposed by K. Simonyan and A. Zisserman in their work. The model was able to achieve 92.7% top-5 test accuracy in ImageNet.

The input to the convolutional neural network during training is a fixed size 224 x 224 RGB image. The only pre-processing we do in the training step is to subtract the mean value of each channel (red, blue, green channels) which was computed on the training set, from each pixel. We pass each image through a stack of convolutional layers and each layer uses a very small receptive field of size 3 x 3. The convolution stride of 1 pixel is used. The spatial padding of the convolutional layer is selected such that the resolution is preserved after convolution. The process of spatial pooling is carried out by max pooling layers which follow some of the convolution layers.

This model is trained in 128X128 images. In this model we have a total of 13 convolution layers and 3 fully connected layers. VGG has smaller filters (3X3) with more depth instead of having large filters. At the end we have the same effective receptive field as if you only have one 7X7 convolution layer.

Residual Networks is a classic neural network(RESNET-50)

ResNet50 is a variant of the ResNet model which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer. It has 3.8 x 10^9 Floating points operations. It is a widely used ResNet model and we have explored ResNet50 architecture in depth.

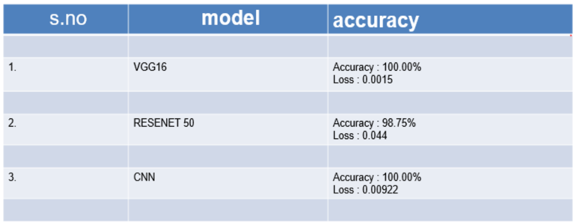
This model consists of 5 stages each with a convolution and identity block. Each of these

convolution blocks has 3 convolution layers and each identity block also has 3 convolution layers. The ResNet-50 has over 23 million trainable parameters.
I have trained this model with the same parameters and number of epochs as CNN and VGG16 so that we can compare the model correctly.

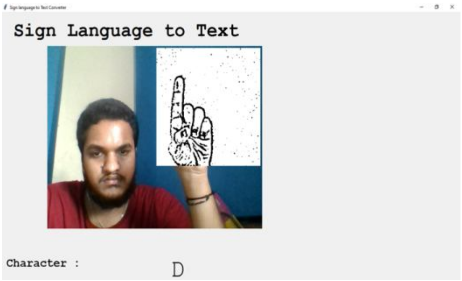
6.RESULT

After analyzing all the tree algorithms VGG-16 gives best accuracy with minimal loss.

Table 1: Model Accuracy



s.no	model	accuracy
1.	VGG16	Accuracy : 100.00% Loss : 0.0015
2.	RESENET 50	Accuracy : 98.75% Loss : 0.044
3.	CNN	Accuracy : 100.00% Loss : 0.00922

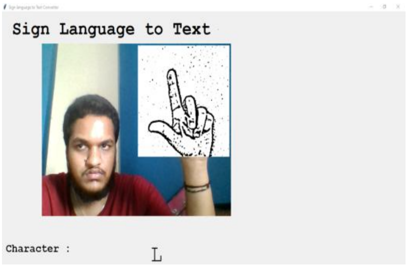


Sign language to Text Converter

Sign Language to Text

Character : D

Figure 4: Sample output-1



Sign Language to Text

Character : L

Figure 5: Sample output-2

7.CONCLUSION

In this project we have built an Alphabet detection system for finger-spell image analysis by building a Vgg-16 based regression system. This system has given us promising results by correctly identifying Alphabets.

REFERNCES

[1] Munib, Q. "Android sign language (ASL) recognition based on Hough transform and neural networks", Expert Systems with Applications, 2000701.

[2] DaviHirafuji Neiva, CleberZanchettin. "Gesture Recognition: a Review Focusing on Sign Language in a Mobile Context", Expert Systems with Application, 2018.

[3] Farid Parvini, Dennis McLeod, Cyrus Shahabi, Bahareh Navai, Baharak Zali, Shahram Ghandeharizadeh,"An Approach to Glove-Based Gesture Recognition".

[4] Maria Eugenia Cabrera*, Juan Manuel Bogado, Leonardo Fermin, Raul Acuña, DimitarRalev, "Glove-Based Gesture Recognition System", Clawar 2012 – Proceedings of the Fifteenth International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines, Baltimore, MD, USA, 23 – 26 July 2012.

[5] HoussemLahiani, Mahmoud Neji, "Hand gesture recognition method based on HOG-LBP features for mobile devices", 22nd International Conference Engineering on KnowledgeBased Systems and Intelligent Information & Engineering System.

[6] Ramesh M. Kagalkar, S.V Gumaste, "Gradient Based Key Frame Extraction for Continuous Indian Sign Language Gesture Recognition and Sentence Formation in Kannada Language:A Comparative Study of Classifiers", JCSE International Journal of Computer Sciences and Engineering.

[7]Wristbands That Translate Sign Language Into Speech Would Be Awesome (gizmodo.com)

[8] Jorge Lobo, Pedro Trindade, "InerTouchHand System - iTH - Demonstration of a Glove Device with Distributed Inertial Sensors and Vibro-tactile Feedback", International Journal of Online and Biomedical Engineering (iJOE) – eISSN: 2626-8493, Vol 9 (2013).

[9] Anita Jadhav, Rohit Asnani, RolanCrasto, OmprasadNilange, AnamolPonkshe, "Gesture RecognitionUsingSupportVectorMachine",InternationalJournalofElectrical,Electronics and Data Communication, ISSN: 2320-2084, Volume-3, Issue-5,May-2015.