Gesture Recognition Using ASL Dataset And Image Analytics

Dr. P. A. Nageswara Rao[1], Ms. G. S. Manjusha[2], Ms. N. Sailaja Reddy[3], Mr. E. Renukanadh[4], Mr. K. Varun Sai[5]
[1, 2, 3, 4, 5] Assistant Professor, Dept of Electronics and Communication Engineering
[1, 2, 3, 4, 5] Gayatri vidya parishad college for Degree and PG courses(A), Visakhapatnam, India

Abstract- *Speech disorder is one of the major disabilities found around the world. The main problem with voice disabled people is the ability to communicate with common people. So, when the disabled people try to communicate their requirements, the common people should also know the sign language. Our project main moto is to build the communication gap between both common people and voice impaired people without using any external components and translator. It can be done by using CNN algorithm, deep learning model, Open CV and ASL dataset. Additionally, we are using Image Analytics to increase the accuracy of the image detection.*

Keywords- Sign language, speech disorder, CNN algorithm, ASL (American Sign Language) dataset, Image Analytics, Open CV.

I. INTRODUCTION

The kaggle repository has ASL image dataset consisting of 3000 images for each alphabet. We took 400 images per alphabet from it and additionally 100 images of our own self generated images with different background and brightness conditions. Totally 14000 images for training the CNN model.

Fig. 1.1 Sample of self-generated Gestures

The above are few gestures for few letters which are self-generated.

II. EXPLORATORY IMAGE ANALYTICSIMAGE

An Image is a combination of pixels. Pixel is a combination of three fundamental/basic colors namely, red, green and blue. Each color will have different intensities for which the values range from 0-255, since, each number is a 8-bit number. Depending on the combination of these three colors with different intensity values gives the different shades. So, there are totally of 16.8 million shades

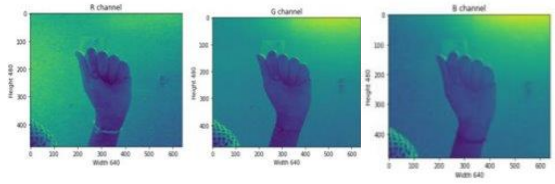


Fig.2.1 Individual sample channels

DISTRIBUTION PLOT TO RGB PIXELS

It gives the distribution plot for the RGB density to the intensity values of a channel at a certain pixel.

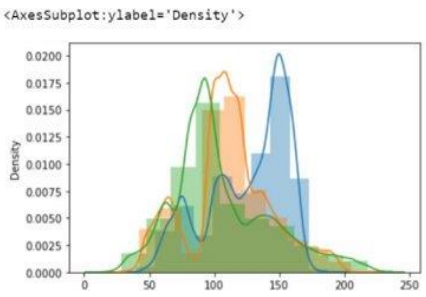


Fig.2.2 Distribution plot to RGB pixels

IMAGE SPLITTING

Image. Split() method is used for the splitting of an image in to individual bands. This will give three images with individual original bands.

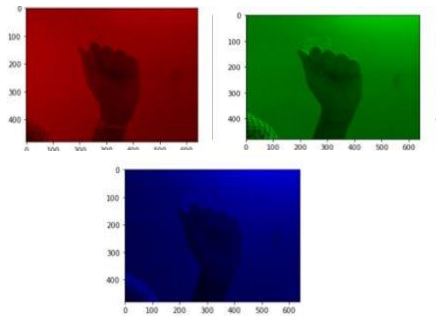


Fig.2.3 An Image split into RGB Channels

IMAGE AVERAGING

It is an image processing technique which is often used to enhance the images. The formula for image averaging is

A (N, x, y) = (1/N) * $\sum_{i=1} I(i, x, y)$

Where 'N' is number of total pixels, 'I' is the intensity of each pixel. The average image is formed by replacing the value with the neighboring pixel value including itself.

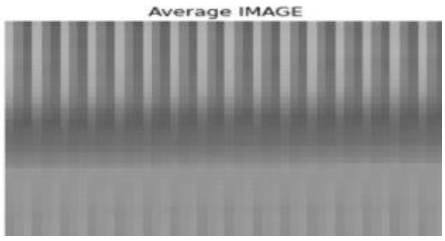


Fig.2.4 The Average of all the Channels in an Image

EIGEN IMAGES

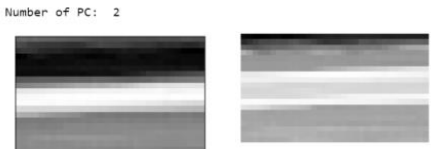


Fig.2.5 The Eigen Image of a sample Image

Eigen images form the basis set of all images used for the construction of covariance matric, which produces the dimension reduction, by allowing the smaller basis set of images to represent the original dataset. This can be achieved by performing PCA (Principal Component Analysis). The eigen images which were created will appear as light and dark areas arranged in a certain pattern. This pattern will be differed from the other gesture as the gesture changes. Initially eigen images are formed by subtracting the average image from the normal image, after which each row of pixels is taken out and formed into vector. As a result, matrix is formed from the extraction of vectors, by which covariance matrix is formed with high probability distribution of vectors and finally, variance matrix of pixels is converted in to image. So, this produces the dimension reduction basis dataset and to represent the smaller basis set to original training images. Real time image recognition is very accurate as it can handle large database and it also reduces the complexity of real time gesture recognition.

EDGE IMAGES

Edge images are formed in the digital images where the brightness changes sharply or where the discontinuities occur. The Prewitt operator is used for the edge detection. The Prewitt operator is used usually in image processing, and mostly for the edge detection. It is a discrete differentiation operator, from computations gives the gradient of the image intensity function. The result of the Prewitt operator at each point in the image is either corresponding gradient vector or norm of this vector. Actually, the Prewitt operator is based on convolving the actual image with a small, separable, and integer valued filter in both horizontal and vertical directions. Prewitt operator used for edge detection in an image, that detects two types of edges: They are:

1. Horizontal edges
2. Vertical Edges

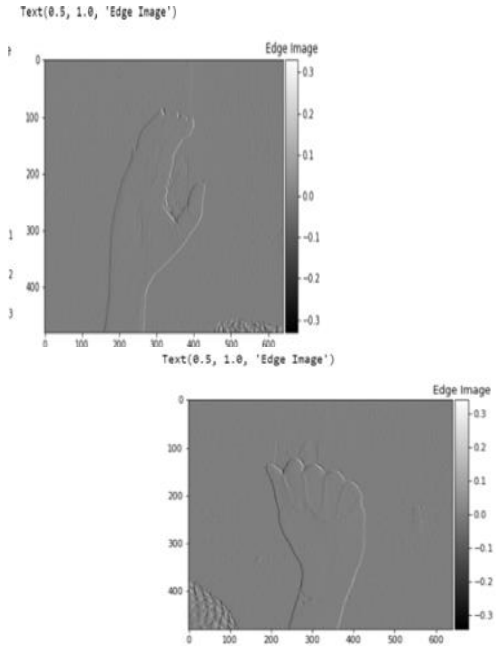


Fig.2.6 The Edge detected image

This operator uses kernels which is of 3x3 to convolve with the original image to calculate the

approximations of derivatives both for vertical and horizontal changes. The operator calculates the gradient of intensities at each pointand gives the direction of possible increase from light to dark and rate of change in that particular direction.

III. DEEP LEARNING MODEL

IMAGE PRE-PROCESSING

The images taken in the dataset are augmented to increase the robustness and accuracy. The list of image augmented techniques is:

1.Rotation by 45 degrees
2. Zoom in and Zoom Out
3. Brightness Range
4. Width and Height Shift
5. Rescaling
6. Resizing to Standard size of 224 X 224

CONVOLUTION NEURAL NETWORK

Convolution Neural Network is a deep learning algorithm that takes input image dataset along with allocating some weights to different aspects in the image. This gives the differentiation from one to the other. As compared to other classification algorithms, the pre-processing required in CNN is much lesser. Filters are hand-engineered in other methods with sufficient training, but in CNN, it has ability to learn the characteristics of filters and build own filters using back-propagation technique. This filter can also be named as "Kernel".

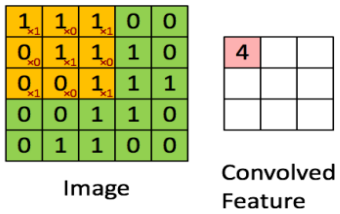


Fig.3.1 The image convolved with A Kernel/Filter

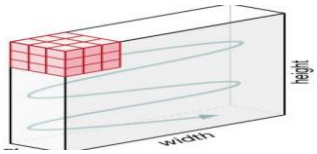


Fig.3.2 Movement of kernel

The yellow-coloured element in the above image is the kernel/filter, that involved in the convolution operation in the first layer of Convolution Layer. The second image represents the movement of the kernel to the right with a certain stride value till it pass through the complete width. In this manner the convolution with kernel/filter takes place with respect to the stride value throughout the whole width. Through back-propagation process, CNN will make its own filter for the convolution process. The first convolution layer is responsible for extracting Low-Level features like edges, colour, gradient, etc., With added layers the architecture adapts/extracts High-Level features, giving us a network that has the wholesome understanding of images in the dataset, similar to how we do.

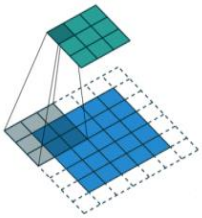


Fig.3.3 Sample convolution operation with Stride Length=2

POOLING LAYER

Pooling layer, like convolution layer, is responsible for the spatial size of the convolved feature. The main aim of pooling is to lessen the computational power required to process the data through dimensional reduction. Apart from that it is also very useful for extracting the dominant features which are rotational and positional invariant, thus maintaining the process of the training of the model effectively.

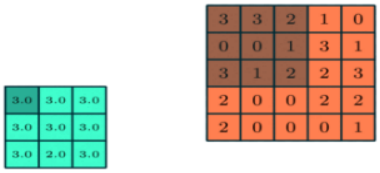


Fig.3.4 Sample of Pooling Layer over the Convolved Layer

Again, pooling is of two types- Maximum pooling and Average pooling.

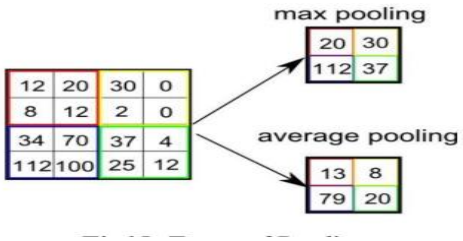


Fig.3.5 Types of Pooling

1. The maximum pooling gives the maximum value of the image portion covered by the kernel.
2. The average pooling gives the average of all the values from the portion of the image covered by the kernel.

ACTIVATION FUNCTION

We have used ReLu (Rectified Linear Unit) in each of the layers, which adds non-linearity to the formula that helps to read complicated features. This activation ReLufunction helps in speeding up the training process by reducing computation time.

Classification — Fully Connected Layer(FCLayer)(The Final Layer)

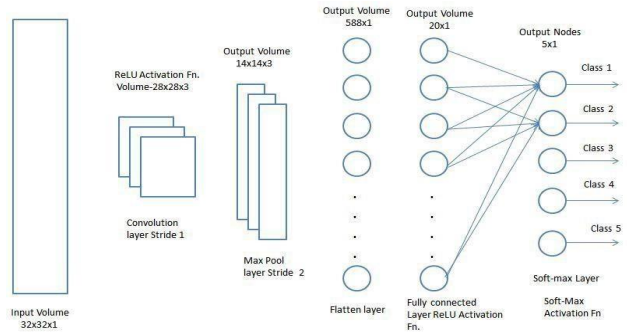


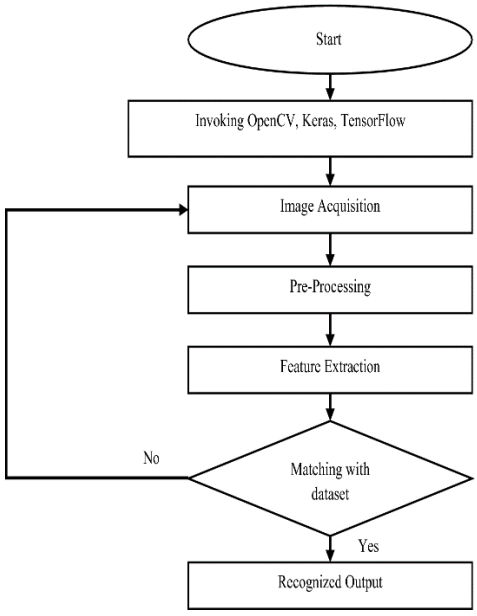
Fig. 3.6 Adding the final layer for classification

The fully-connected layer is used for finding the possibly non-linear function in that space. The image should be flattened column vector. The output which was flattened will be fed to the forward neural network and back propagation is applied at every iteration of training. After a series of epochs, the model will have the capability to distinguish between dominating and Low-Level features in images and will classify them using SoftMax Classification technique.

OpenCV

OpenCV(Open-Source Computer Vision Library) is a library of Python bindings which is designed to solve computer vision problems. It is also a great tool for image processing tasks. It also has access to multiple languages along with python, which are JAVA and C++. As it is an open-source library, it is quite easy to perform machine learning tasks. By using open CV tasks like object detection, Face detection, etc., can be performed.

OUTPUT ACQUISITION FLOW



IV. SAMPLES OF RESULTS OBTAINED

Fig. 4.1 Gesture for I

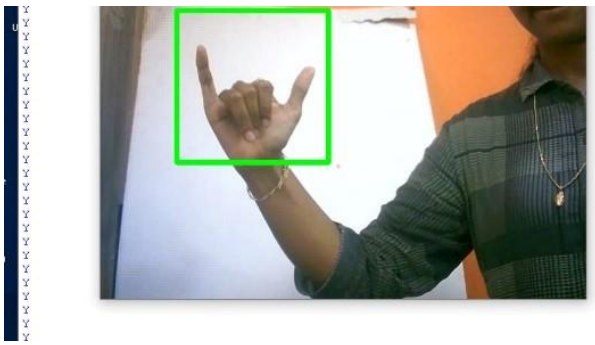


Fig. 4.2 Gesture for letter Y

V. CONCLUSION

Inability to speak and hear is one major challenge for the human race. As an easy and practical way to achieve human-computer- interaction, this solution of gesture to speech and text conversion has been used to facilitate the reduction of hardware components. This solution aims to provide aid to those in need thus ensuring social relevance. By this, the people can easily communicate with each other. The

application is also cost efficient and eliminates the usage of expensive technology. Hence, Hand Gesture recognition was successfully implemented and the implementation of this system gives up to 93% accuracy and works successfully. As an easy and practical way to achieve human computer-interaction, this solution of gesture to speech and text conversion has been used to facilitate the reduction of hardware components.

VI. FUTURE WORK

Future improvements to our framework may be to expand the application to a wider vocabulary. Since there are many other sign languages (BSL etc.) it can be extended to several gestures according to the different sign languages. Further, more advanced algorithms may replace the current in place so as to improve the accuracy and processing speed of the system. The same can be used for object detection for classifying different objects. And also, text to speech can also be included.

REFERENCES

[1] S. Patel, U. Dhar, S. Gangwani, R. Lad, and P. Ahire proposed a system title named Hand-gesture recognition for automated speech generation, 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT).

[2] ItkarkarRajeshri R, Anil Kumar V. Nandi, Vaishali B. Mungurwadi Indian Sign Language Recognition using Gabor Feature extraction, 2019 International Journal Of Scientific & Technology Research 4 Vol 8, Issue 11, NOVEMBER 2019 ISSN 2277-8616 .

[3] Shuhangi Shinde, R RItkarkar ,Anilkumar Nandi Gesture to Speech Conversion for Sign Language Recognition, 2017 International Journal of Electronics, Electrical and Computational System ISSN 2348- 117X Volume 6, Issue 8 August 2017.

[4] Sonali Munde1 , Prof. RajashriItkarkar, Dr. D.G. Bhalke, Dr. D. S. Bormane Automatic Detection of Text Labels of Hand-Held Objects for Blind persons, 2017 International Journal of Electronics, Electrical and Computational System ISSN 2348- 117X Volume 6, Issue 8 August 2017.

[5] Shilpa Bhople, ItkarkarRajashri, Pooja Bhoir Indian Sign Language Recognition, 2015 International Journal of Electrical, Electronics and Computer Science Engineering , Volume 2, Issue 3 (June 2015)| E-ISSN : 2348-2273|PISSN:2454- 1222.

[6] S. Vigneshwaran, M. Shifa Fathima, V. Vijay Sagar, and R. Sree Arshika Hand Gesture Recognition and Voice Conversion System for Dump People, 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS).

[7] P. Vijayalakshmi and M. Aarthi "Sign language to speech conversion", 2016 International Conference on Recent Trends in Information Technology (ICRTIT), pp. 1-6, April 2016.