International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

Sign Language Interpreter using Deep Learning

Rachaell Nihalaani[1], Apoorva Shete[2], Vaibhav Ambhire[3]
[1]Computer Engineering Department, Thadomal Shahani Engineering College
[2]Electronics and Telecommunications Department, Thadomal Shahani Engineering College

Abstract: Sign Language is invaluable to hearing and speaking impaired people and is their only way of communicating among themselves. However, it has limitations with its reach as the rest of the people have no information regarding sign language interpretation. Sign language is communicated via hand gestures and visual modes and is therefore used by hearing and speaking impaired people to intercommunicate. These languages have alphabets and grammar of their own, which cannot be understood by people who have no knowledge about the specific symbols and rules. Thus, it has become essential for everyone to interpret, understand and communicate via sign language to overcome and alleviate the barriers of speech and communication. This can be tackled with the help of machine learning. This model is a Sign Language Interpreter that uses a dataset of images and interprets the sign language alphabets and sentences with 90.9% accuracy. For this paper, we have used an ASL (American Sign Language) Alphabet. We have used the CNN algorithm for this project. This paper ends with a summary of the model's viability and its usefulness for interpretation of Sign Language.
Keywords: Sign Language, Machine Learning, Interpretation model, Convoluted Neural Networks, American Sign Language

I. INTRODUCTION

According to recent statistics, there are around 700,000 to 900,000 speech impaired people in the world. Of these, 63% are born deaf whereas the others lose their hearing due different factors like accidents, etc. Sign Language is the means of communication used by speech impaired people. With the help of visual means and hand gestures, people who cannot speak can get their thoughts across to the other person. The set of rules, grammar and alphabets are different for all the sign languages. It is not easy for everyone to understand sign languages unless they know the gesture for every alphabet. There are several standardized sign languages like the Indo-Pakistan Sign Language, British Sign Language (BSL), American Sign Language (ASL), etc. The gestures and symbols used in each of these are greatly varied. It is not possible for everyone to know each of these languages. A sign language Interpreter can be used for this purpose.

Sign Language Interpreter can be utilised for interpreting and understanding sign languages. It uses a set of images and hand gestures to interpret what a person is trying to communicate via sign language and translates these hand gestures into normal English alphabets. This can reduce the difficulties faced by speech impaired people while communicating their thoughts and ideas, to a large extent. Due to the important factors that can help various people communicate efficiently with each other, demands for such Sign Language Interpreters are increasing day by day. Such systems give the opportunity to the speech impaired people to communicate with others easily. The necessity for such an efficient model that can interpret different sign languages is increasing every day and is becoming increasingly vital. This paper attempts to remove the barriers to communication faced by the speech impaired people with the help of efficient ML techniques. For training our model, the CNN algorithm is used. This paper utilises a collection of varied images for BSL and ASL for interpreting the sign languages and translating them to normal english language with good accuracy and results. This paper is structured in the following manner. Section 2 discusses the relevant research done in the past for interpretation of sign language using ML. Section 3 gives a detailed description of the dataset used and created for building the sign language interpreter. In Section 4, the methodology followed for the implementation of this model is discussed in detail. The following section 5 discusses the results of our experiment, and the conclusion and ways in which this model can be improved in the future are discussed in the last section of this paper.

II. LITERATURE REVIEW

In this section, the relevant research and the previously implemented sign language interpreters are discussed.

The paper [1] implements an application that uses a webcam for recognizing hand gestures for ASL and converting these signs and gestures to text. This translated text is further converted into audio thus providing a sign language translator. This paper uses the Convoluted Neural Network (CNN) algorithm for interpreting and converting hand gestures to text. The accuracy achieved is 95%. In paper [2] the main objective is to build a system that recognizes the sign language gestures and converts them to corresponding texts.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

Here too, the hand gestures are captured by the computer webcam and using the pose estimation library the images are processed. The data obtained from the processed images is then converted to a csv file and this is the dataset used for building the sign language interpreter. This paper uses the Decision Tree algorithm for training and testing the implemented model. In [3], the paper uses CNN algorithm for implementing a fingerspelling translator. This paper utilizes the GoogLeNet architecture and two distinctly different ASL datasets for the training and testing of their model. The accuracy obtained were: 98% validation accuracy with five letters and 70% with ten. The paper [4] implements an Indian Sign Language Interpreter using different ML algorithms. The Gesture Area Acquisition Module is used for recognizing and locating hand gestures while ignoring the background factors, blur, etc. Paper [5] also uses the CNN algorithm for interpreting the signs of Swedish Sign Language (SSL) and consists of a pre-trained Inception V3 Network. The system accuracy achieved in this paper is 80%, based on 8 study subjects and 9,400 different sets of images. A detailed summary of the methodology followed, the results obtained and the future scope of this model is also given in the paper. The paper [6] creates an application that captures the sign language from the computer's webcam, the image data is then processed using a combinatorial algorithm and is recognized using template matching. A data of 6000 images is used of which 4,800 images are used for model training and the remaining 1,200 images are used for testing the model. Indian Sign Language (ISL) is translated using the SVM algorithm with 88% system accuracy. In [7], the Arabic Sign Language (ArSL) and ASL translated to both text and audio using both static and dynamic gestures. The accuracy achieved for static gestures is 95% and for the dynamic gestures it is 88%. This literature review and knowledge about the past research in this field provided us with deep insights and gave us a pathway to go about our research.

III. DATASETS

Dataset collection is the first and foremost step. This sign language interpreter requires two datasets. The first is the ASL Alphabet [8] dataset obtained from Kaggle. This dataset contains 3000 image files for each of the 26 alphabets, and each of the three additional categories - delete, space and nothing. This constitutes a total of 87000 images. The second dataset i.e. the Keras Pretrained Model dataset [9] helps us use a pre trained keras model in our interpreter. Specifically, we have used the VGG16 convolutional neural network model that is a part of this dataset.

IV. METHODOLOGY

The Sign Language Interpreter model was developed in Python language. It made use of the open source python library Keras CNN that acts as an interface for the TensorFlow library as well as for artificial neural networks. The following is a brief outline followed in developing this model. First, both the datasets - the ASL Alphabet dataset containing images [8] and the Keras Pretrained Models Dataset [9] - were loaded. This data was then explored, pre-processed and visualised. Pre-processing of data involves assigning labels to images such as 0 for images with the sign representation of alphabet 'A' and so on. In total, we have 30 such labels for images including all the alphabets (A-Z), the delete symbol, the space symbol and nothing. These labels are then encoded into hot vectors. A hot vector is a 1xN matrix used to uniquely identify the label. After shuffling to permit further subsampling, the dataset is split in a 80:20 ratio, into training and testing sets respectively. Now the dataset is ready to be used. To familiarise ourselves with the dataset, we plot a few images from category 'A'. To gain a better understanding of the data, we use the visualisation technique of plotting histograms of RGB Pixel Intensities.

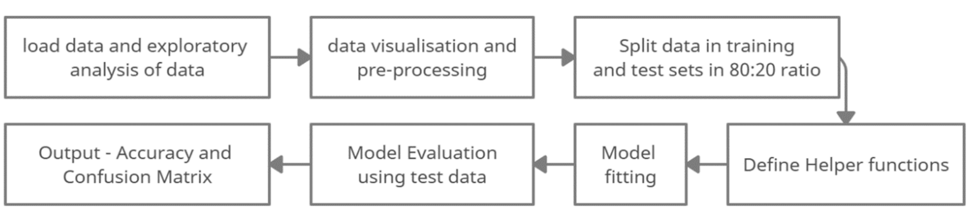

Fig. 1 Proposed Methodology Block Diagram

Now that we are ready to actually develop the model, we define a few helper functions that we will require to fit and evaluate our model. These include ones that aid in plotting confusion matrices and learning curves. Next, we evaluate the model. We have implemented the use of transfer learning with VGG16. Transfer learning entails reusing a pre-trained model on a new but related problem, and with comparatively less data. VGG16 is a convolutional neural network model proposed in [10]. This brings us to why the second dataset [9] was required, which is because it contains the VGG16 deep learning model. The model is fit over the training data and then evaluated over the test data.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

V. RESULTS AND ANALYSIS

Careful observation of this model on two datasets, [8] and [9], have given us the following results illustrated in this section. Before we delve into details of the results, we must familiarise ourselves with the ASL Alphabet dataset [8]. Fig. 2 shows 20 images from category 'A' that show the sign language gesture representing the letter 'A'.

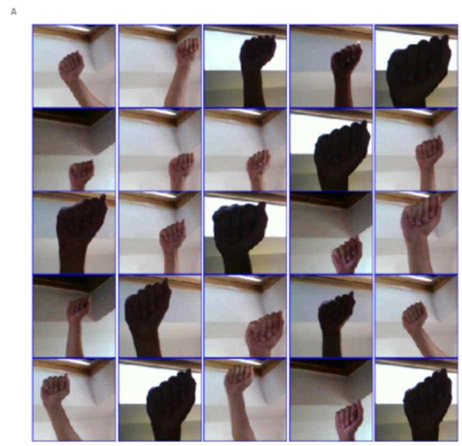

Fig. 2 Few instances of dataset – alphabet 'A'

We have also used the visualisation technique of plotting histograms of RGB pixel intensities in order to gain a deeper understanding of how the images are interpreted by our model.

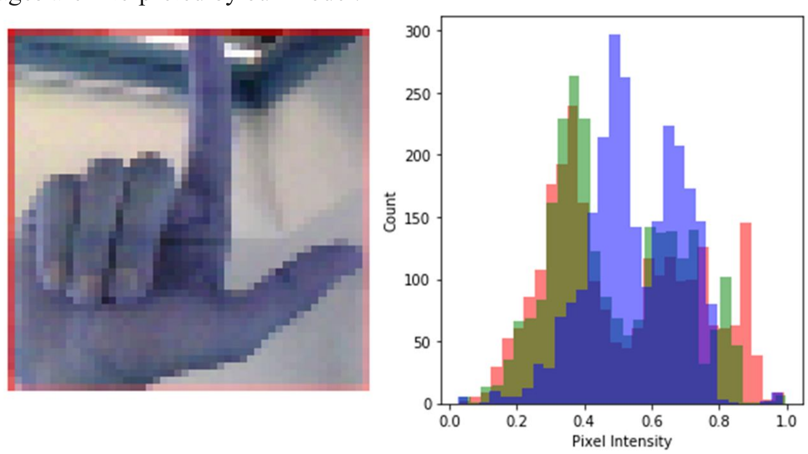

Fig. 3 Image Visualisation using Histograms

As for the final results, our Sign Language Interpreter has shown evaluation metric values as in Table 1.

TABLE I
Sign Language Interpreter Results

Performance Evaluation Metric	Value
Accuracy	0.90
Precision	0.91
Recall	0.91
F1 Score	0.91

The above indicate all average values. All of these metrics have been computed for each of the 29 labels. These model evaluation parameters are defined as follows. Accuracy is the portion of correctly classified data instances. Precision is the true positives percent of predictive positives. Recall is the true positives percent of actual positives. F1 Score is the harmonic mean of precision and recall. [11]

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

A confusion matrix is a table that describes model performance. Fig. 4 shows the confusion matrix we obtained.

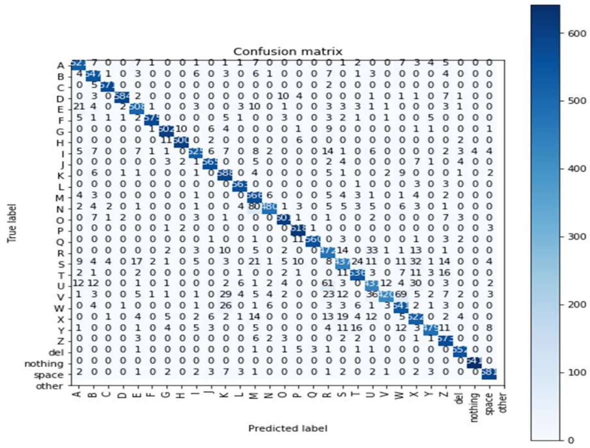

Fig. 4 Confusion Matrix

VI. CONCLUSION AND FUTURE SCOPE

In today's advanced and accommodating world, there have been ample opportunities for persons with disabilities to succeed. This sign language interpreter aims to further bridge the communication gap that they face. Ideally, everyone should be considerate enough to learn sign language in order to enable persons with disabilities to feel as equals. However, considering that there is a long way to go before that can happen, in the meantime, machine learning can be a powerful tool in enabling all parties involved. Our model has shown a 90% accuracy, which is a commendable place to begin. As is with machine learning in general, our model can also have scope to improve. For instance, this sign language interpreter can be trained over various other kinds of sign language datasets such as BSL (British Sign Language) and ISL (Indian Sign Language). This enables more people, from all over the world, to benefit from this model. Further, other deep learning models and network architectures can be applied with the aim of achieving better accuracy. The denouement of this sign language interpreter would be an application interface that reads gestures in real time and aids in communication.

VII. ACKNOWLEDGMENT

We would like to express our gratitude to Professor Vaibhav Ambhire for his immense support and guidance. We would also like to express our gratitude to our Principal, Dr. G.T. Thampi.

REFERENCES

[1] A. Ojha, A. Pandey, et.al. (2020). Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network. Presented at NCAIT - 2020 Conference Proceedings. [Online]. Available: https://www.ijert.org/research/sign-language-to-text-and-speech-translation-in-real-time-using-convolutional-neural-network-IJERTCONV8IS15042.pdf

[2] H. Gowda, V. Chandra, S. Vishwas, et. al. (2018). Sign Language Translator Using Machine Learning. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 4 (2018) Spl. © Research India Publications.[Online]. Available: https://www.ripublication.com/ijaerspl2018/ijaerv13n4spl_01.pdf

[3] B. Gracia, S. A. Viesca. Real-time American Sign Language Recognition with Convolutional Neural Networks. [Online]. Available: http://cs231n.stanford.edu/reports/2016/pdfs/214_Report.pdf

[4] A. Chavan, A. Bhat, S. Mishra, et. al. (2021). Indian Sign Language Interpreter for Deaf and Mute People. www.ijcrt.org © 2021 IJCRT | Volume 9, Issue 3 March 2021 | ISSN: 2320-2882. [Online]. Available: https://www.ijcrt.org/papers/IJCRT2103178.pdf

[5] Halvardsson, G., Peterson, J., Soto-Valero, C. et al. Interpretation of Swedish Sign Language Using Convolutional Neural Networks and Transfer Learning. SN COMPUT. SCI.2, 207 (2021). https://doi.org/10.1007/s42979-021-00612-w

[6] O. Vedak, P. Zavre, A. Todkar, et. al. (2019). Sign Language Interpreter using Image Processing and Machine Learning. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072. [Online]. Available: https://www.irjet.net/archives/V6/i4/IRJET-V6I4413.pdf

[7] S. A. E. El-Din and M. A. A. El-Ghany, "Sign Language Interpreter System: An alternative system for machine learning," 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020, pp. 332-337, doi: 10.1109/NILES50944.2020.9257958. https://ieeexplore.ieee.org/document/9257958

[8] ASL Alphabet Dataset. Kaggle.com. Available:https://www.kaggle.com/grassknoted/asl-alphabet

[9] Keras Pretrained Model Dataset. Kaggle.com. Available:https://www.kaggle.com/gaborfodor/keras-pretrained-models?select=imagenet_class_index.json

[10] S. Simonyan, A. Zisserman. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. Arxiv.org. [Online]. Available: https://arxiv.org/abs/1409.1556

[11] Harikrishnan N B. "Confusion Matrix, Accuracy, Precision, Recall, F1 Score." medium.com. https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd (accessed Sep. 10, 2021).