Sign Language Recognition Using Inception V3 Model

Akshay Vijayan[1], Neevan R[2]
[1, 2]Dept of Computer Application
[1, 2]College of Engineering, Trivandrum, Kerala, India

Abstract- *Gesture recognition has been a research area since the emerging of machine learning and image processing. Sign Languages are the way through which the deaf community interact with each other. For a person who doesn't know sign language find it difficult to communicate with deaf people as the are not able to understand the sign language. This paper proposes a inception v3 model based sign language recognition system which is capable of recognise sign languages from images and live video. As the model uses the most effective algorithm for multi-classification problem it achieves a fair test accuracy of 96%.*

Keywords- Sign Language Recognition, Inception V3, ConvNet, Image Processing, Pattern Recognition.

I. INTRODUCTION

Sign Language is one of the most discussed subject in assistive technology development. It is a language with its own grammar which is easily understandable. The deaf people use sign languages for communicating easily than any other methods. Many innovations is happening in the area of assistive technologies and sign language recognition is one of the important area of research as the developers or the researchers are not able to develop a fully confident system. The main issue in the recognition process is the number of classes for classification. The basic character set has 26 classes which is the 26 alphabets in English.

The multiclass classification problems needs a huge amount of data as there should be sufficient features to differentiate each class of data. Feature extraction and tuning plays the soul of the architecture modelling of a system. Here that process is carried out with care to design a efficient sign language recognition system. Timely evaluation on processing and testing are carried out with visual analysis.

II. SYSTEM OVERVIEW

There are a lot of Sign Languages across the worlds such as French Sign Language, Indian Sign Language, American Sign Language etc. Almost all the sign languages process the same lexicon. The most widely used sign language is the American Sign Languages which uses the single hand for each alphabets and which makes it more convenient for people to communicate. In the paper American Sign Language Dataset is used which is collected from different portals.

Fig 1. American Sign Language

In case of recognition few letters makes the system complex as the letters are almost similar in shape. It can be seen in the case of alphabet "A" and "S", there is only a slight change in the feature. So if the system needs to differentiate it fastly then sufficient of data should be supplied for these complex signs.

The current dataset contains almost 200 images per class. Model developed from this dataset can only be used as a proof of concept as more amount of data is needed for a accurate model. The inception v3 model architecture overcome the issue of less of data in each class as it is the best model for multi classification.

III. TECHNOLOGY STACK

A. DEEP LEARNING

Deep learning is a machine learning strategy where the features are automatically selected extracted. It differs from the traditional neural networks due to the presence of multiple hidden layers. The number of hidden layers is a critical factor in deciding the optimality of the system.

The architecture of the model plays a prominent role in determining the feasibility of the system. We have developed a new model based on ConvNet which is a deep learning algorithm.

B. TENSORFLOW

Tensorflow is a framework designed for machine learning model development. It helps in the easy processing of machine learning algorithms. It has dynamic methodologies for performance evaluation. It provides API to incorporate in different modules to get machine learning models possible in all the application interface.

C. OPEN CV

It is an image processing tool used in the feature extraction of images. In this paper the use of OpenCV is in live video recognition. It is the best open source tool for live video-image processing.

D. TKINTER

Tkinter is a python application development tool. This helps to develop interfaces for applications that looks like a software UI. Tkinter can only be used with the applications developed in python as Tkinter is a python application interface development tool.

IV. SYSTEM DESIGN

A. TRANSFER LEARNING

The latest recognition systems needs a lot of parameters and arguments to design a good model. Developing a model from scratch demands a lot computation power as the training may take a lot of days. It affects the bug fixing as we have to wait for days to do bug fixing and model tuning. So to overcome this issue the concept of Transfer learning is used here. It is a methodology where the pre-learned material will be used to develop another application on the top of it. Transfer learning strategies are widely used as it is revolutionising the field of computational efficiency in the model development. Re-purposing of existing general system is the core of the mechanism proposed in this paper. As shown in the Figure 2, the model with transfer learning has a higher performance.

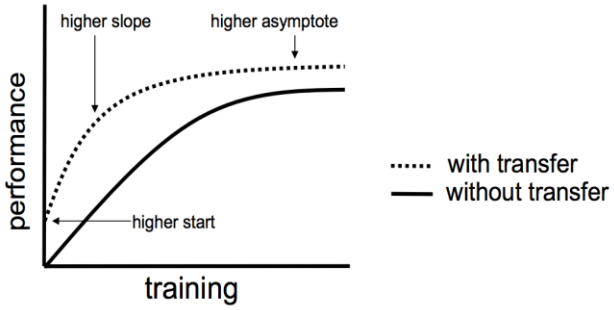

Fig 2. Transfer Learning

B. INCEPTION V3 MODEL

Inception V3 is a architecture designed by Google which is capable of classifying even 1000 classes of data. The error rate of Inception model is less that the human error rate. The model is trained on ImageNet dataset which contains more than 2400 data classes. ImageNet is one of the largest image dataset. As the model is trained on this dataset, maximum number of feature identification layers are effectively added in the inception architecture as shown in Figure 3.

The model also has all the operations in a ConvNet architecture including Convolution, Pooling, Drop out. The convolution operation is done to join two feature set to generate a feature map. The average and max pooling are done to get the best features from the data. It neglects features with less importance through a sliding pooling window.

This pre-trained model is re-trained with our sign language image dataset on the top layer where the final part is 8 x 8x 2048 size. The last fully connected layer and the softmax are the layers where the changes are done and all other layers will be intact. Retraining steps are having more codebase as compared to a normal ConvNet architecture. To train a model similar to Inception in a normal CPU system may take around a year.

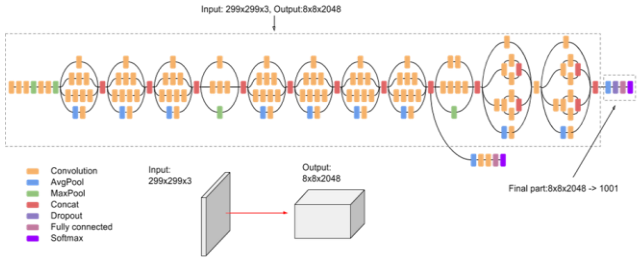

Fig 3. Inception V3 Model

Inception model has to be downloaded to the local system before retraining starts. The feature extraction is done with the downloaded inception model.

On re-training bottlenecks are created for each image in each class. The bottleneck files will be created which will be a file with numerical values representing the features.

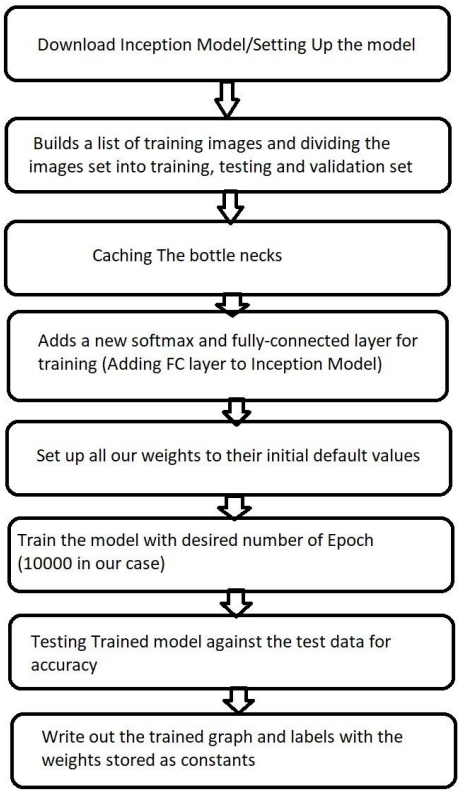

Fig 4 . Algorithm

A new softmax fully connected layer is added for training. The training results in a Neural network weight and labels which are stored as graphs and numerical data files. The complete working algorithm is illustrated in the Figure 4.

Initially the weights will be set to defaults. This model uses backpropagation for weight determination.

C. PREDICTION

The system has a user interface which is designed using Tkinter. The user will be able to browse the image and after uploading the image the system will show the predicted sign character on the application interface itself with probability score for each letter. Also a live video based prediction with confidence score is also there.

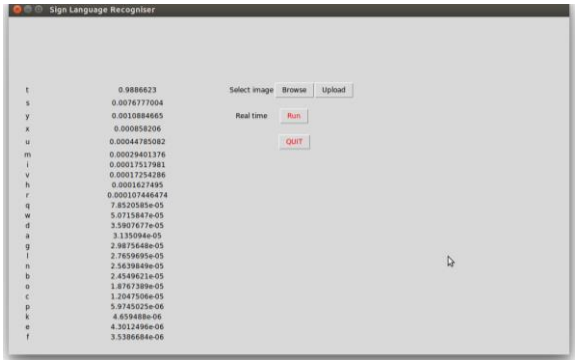

Fig 5. Interface in Tkinter

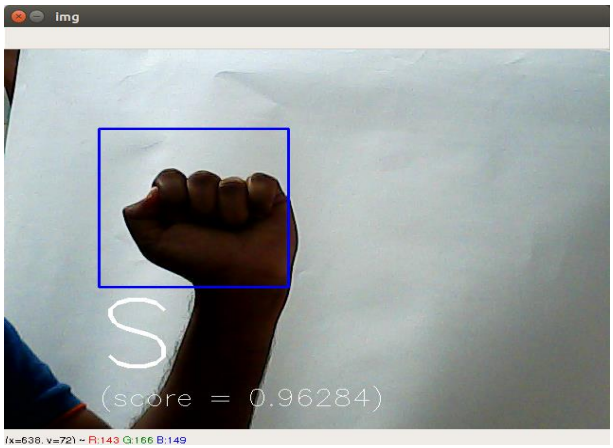

Fig 6. Prediction from Live Video

E. RESULTS

The results are evaluated with the help of the tool Tensorboard. It helps in the visualisation of entropy and all other parameters for evaluation. The model achieved an test accuracy of 96% and validation accuracy of 100%.

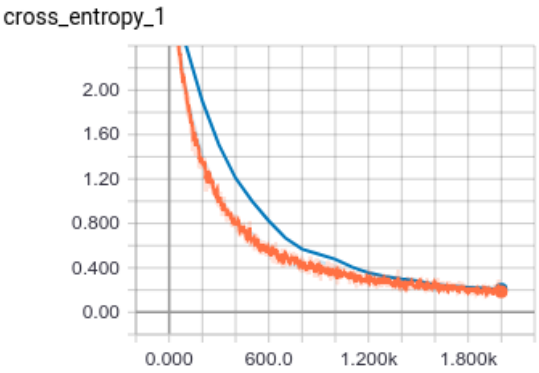

Fig 7. Cross Entropy Rate

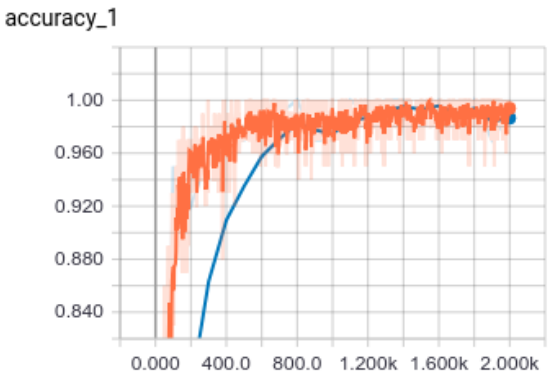

Fig 8. Visualisation of Train & Validation Accuracy

V. CONCLUSION

Assistive technology being the one of the main problem statement scenario in developing technology, it is very important to address the issues with a proper system. This paper proposes a efficient solution for Sign Language Recognition where the complexity is overridden with a high-perform model. The following conclusions can be deduced from the development of the project: It provides an abstract platform between the user and avoid manual work. CNN inception v3 model has improved the efficiency as discussed before.

As the evaluation shows that the model is a good one with a fair accuracy , this can be further developed to a bigger model with a large amount of data. Reducing false positive rate can be taken as a major point for further development.

VI. ACKNOWLEDGMENT

We are grateful to the Almighty for the good health and well being that were necessary to complete this work. We would also like to express our sincere gratitude to Prof. Jose T Joseph, Head of Department and Prof. Baby Syla L, Department of Computer Applications, College of Engineering Trivandrum imparting their knowledge with us and for all the help and support they have given us from the start of the work till its completion.

REFERENCES

[1] Ashish S. Nikam, Aarti G. Ambekar, "Sign language recognition using image based hand gesture recognition techniques", *IEEE Xplore. 2016 Online International Conference on Green Engineering and Technologies*.

[2] D. Aryanie, Y. Heryadi, "American Sign Language-Based Finger-spelling Recognition using k-Nearest Neighbors Classifier", 3rd International Conference on Information and Communication Technology (2015) 533-536.

[3] T.Starner and A. Pentland., "Real-Time American Sign Language Recognition from Video Using Hidden Markov Models", *Computational Imaging and Vision, 9(1); 227-243, 1997.*

[4] J. Atwood, M. Eicholtz, and J. Farrell, "American Sign Language Recognition System. Artificial Intelligence and Machine Learning for Engineering Design", *Dept. of Mechanical Engineering, Carnegie Mellon University, 2012.*

[5] L. Pigou et al., "Sign Language Recognition Using Convolutional Neural Networks", *European Conference on Computer Vision 6-12 September 2014.*

[6] K. Oyebade, O. Ebenezeri, A. Khashman, "Deep Learning in Character Recognition Considering Pattern Invariance Constraints", *International Journal of Intelligent Systems and Applications*, vol. 7, no. 7, pp. 1, 2015.

[7] Y. Jia, "Caffe: An open source convolutional architecture for fast feature embedding", *http://caffe.berkeleyvision.org/, 2014.*