SRI CHANDRASEKHARENDRA SARASWATHI VISWA MAHAVIDHYALAYA
(Deemed to be university u/s 3 of UGC act 1956)
(Accredited with "A" by NAAC)
Enathur, Kanchipuram – 631561. Tamilnadu
www.kanchiuniv.ac.in

Automated Hand Gesture Recognition model

Guided By:
Dr. N.Kumaran
Assistant Professor
Dept. of CSE

Final Review
Date: 15-04-2021

Team Members:

M Sampath (11179A146)
M Sri Anurag (11179A148)

Contents

Abstract

Sign language is an ancient way of non-verbal communication used by the deaf and mute community. It uses hand gestures as a mode of communication for interacting and conveying ideas visually. Recognition of hand gestures is a popular field of research where researchers are implementing various kinds of techniques and trying to commercialize their models, well every researcher has their way and methods to represent their models by improving the limitations of previous models. Here we propose a simple and unique way of hand gesture recognition of American Sign Language. Our model takes threshold images as input to train the model which will help to overcome variations in different skin colors and perform the prediction. We secure a 99.96% of accuracy on training our model using Convolutional Neural Network and achieved a satisfying result on 26 fingerspellings. Overall, this project focus on Automated Hand Gesture Recognition using transfer learning techniques.

Keywords: American Sign Language; convolutional neural network; hand gesture recognition; transfer learning.

Introduction

Hand gesture is a symbol of physical behavior or emotional expression. The main aim of this project is to convert hand gestures to text and further in to speech by helping the people with disabilities like deaf and mute (D&M) to convey their thoughts or ideas to other persons who are unaware of visual language. Sign language is the most widely used visual-manual modality to convey meaning. It's the main form of communication for the Deaf and Hard-of-Hearing community, but sign language can be useful for other groups of people as well. People with disabilities including Autism, Apraxia of speech, Cerebral Palsy, and Down syndrome may also find sign language beneficial for communicating. According to the World Federation of the Deaf (WFD), there are around 72 million people worldwide who use sign language. There are many varieties of sign languages developed naturally through different groups of people, there are somewhere between 138 and 300 different types of sign languages used around the world. Here we focus on recognition of **fingerspellings** in American Sign language using state-of-art algorithms of Convolutional Neural Network by transfer learning techniques.

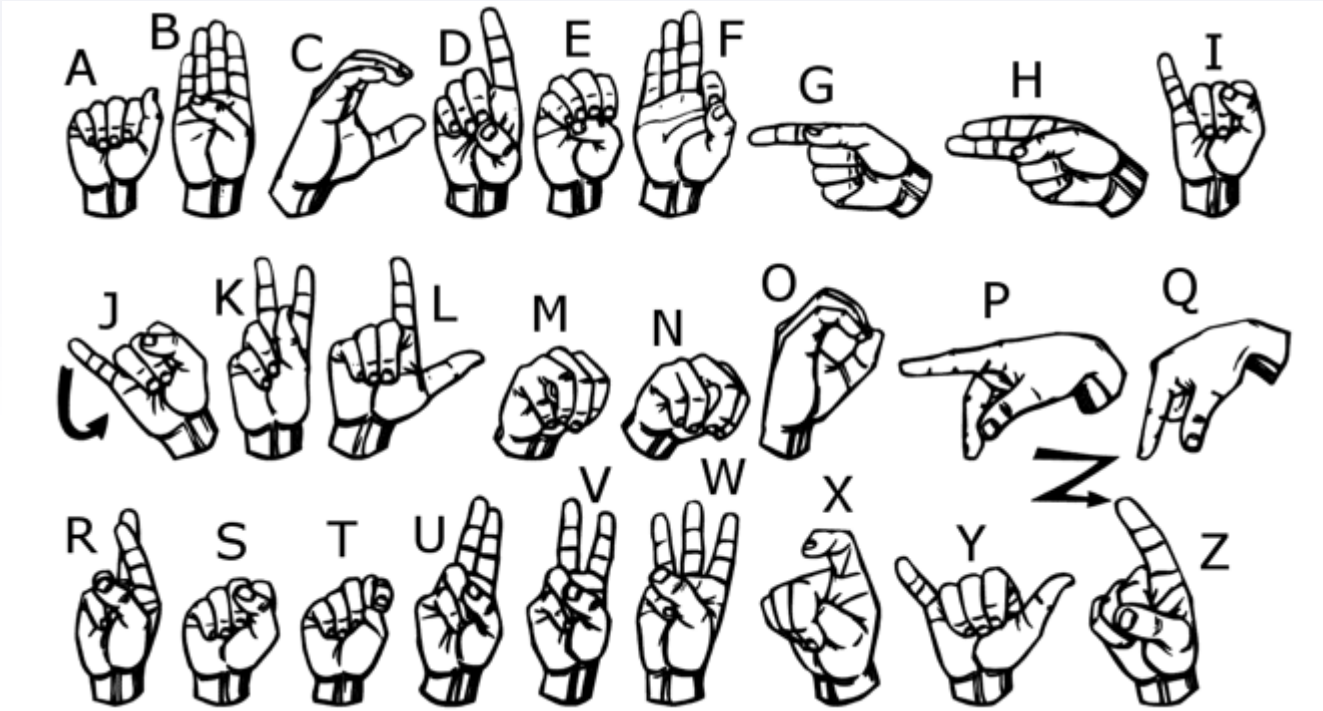


Fig.1: Fingerspellings in American Sign Language

Fig.2: Threshold Image used in existing system, Reference [2]

Existing System

The Hand gesture recognition is a interesting field of research and many models have been proposed by researchers and developers. Some are them are highly advanced while some of them are primitive. The existing systems have different approaches like Smart gloves for hand gesture recognition, hand gesture recognition using binary images, using skin color based approach, using Machine Learning Techniques, Deep learning approaches and many other. The model proposed in [2] is detecting hand gesture using hand segmentation techniques, which include sample of skin color for detection of hand which could be a drawback of using model with different skin colors of hand and requires a static background for prediction. In paper [3] we can observe of using sensor devices for predicting the hand gesture which can be expensive. Similarly there are many models proposed which use different techniques for achieving different set of results.

Proposed Work

Accuracy: The accuracy of the proposed model is comparatively high than existing models i.e. 99.96% of accuracy.

Independent of skin color: Unlike other models our model doesn't depend on skin color of user. It is added advantage as if model encounters various skin color it would reduce efficiency of model.

Faster & Stable Prediction: The proposed model is able to predict the output gesture faster and with more stable than the existing system.

More efficient and less expensive: Our model don't require any additional equipment, it is done using local device web camera and is more efficient when compared to previous models.

Easy adaptability: Our model can easily adapt new gestures and is efficient enough to give higher accuracy results even with many no.of classes.

Convolutional Neural Network

- **Neural Networks** are the most efficient way for replication of the human brain, these neural networks act as neurons and can learn at a very deep extinct.
- A **Convolutional Neural Network (ConvNet/CNN)** is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.
- As Convolutional neural network play a significant role in classification of images, our model use CNN for recognition of hand gestures to achieve adequate results.
- Using of CNN can also help in reducing feature extraction process as the network learns to extract features while training.

Transfer Learning

Transfer learning is the idea of overcoming the isolated learning paradigm and utilizing knowledge acquired for one task to solve related ones. In simple world it is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned.

We used transfer learning to train or model as it helps in reducing the computational cost by utilizing the parameters and features of **pre-trained** models. We trained our model with various pre-trained models of CNN like VGG16, VGG19, Resnet50. we have also performed the training of model using Sequential model of CNN which is **a plain stack of layers** where each layer has **exactly one input tensor and one output tensor.**

Structure of model

Input Gesture → Image Pre-processing → Model Training → Image Classification → Output

- The first step is collecting input gestures from user and applying pre-processing techniques.
- We performed the **Adaptive Gaussian Thresholding with Otsu's Binarization** to completely reduce the noise in the image. The final output of image after preprocessing is shown below in Fig.3.

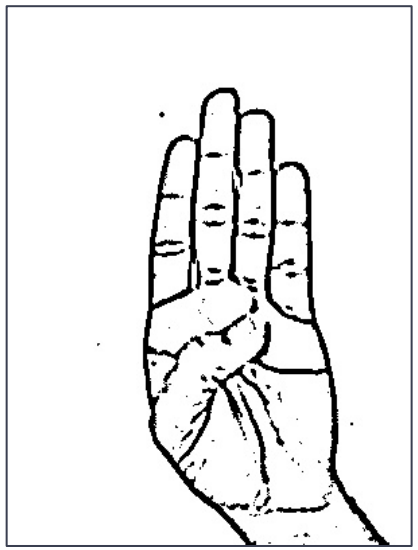


Fig.3: Threshold Image

- This Threshold image is fed as input to the neural network for training the model. After successfully training the model the weights are saved and performed prediction by loading the model with help of TensorFlow and Keras.
- The gestures classification performed in this project are dynamic and provide an satisfying results on 27 classes i.e. 26 fingerspellings of American Sign language and additional 1 class for blank image, with an **accuracy of 99.96%** and with negligible **loss of 0.00114**. Sample output of the model are given below in the Fig.4 (a), (b).

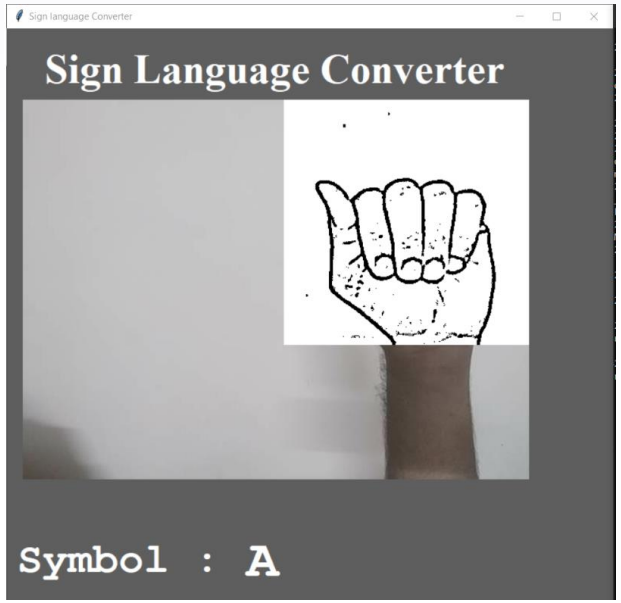


Fig.4 (a): Final output of model predicting fingerspelling “A”

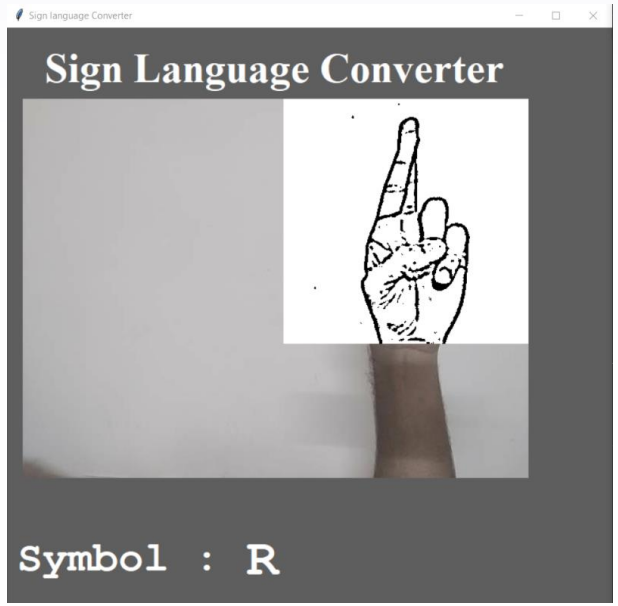


Fig.4 (b): Final output of model predicting fingerspelling “R”

Tested Environment

We have performed the training of model using various pre-trained models and even using a sequential model by defining own set of layers. The results obtained by each model are satisfying and there is difference in accuracies of each model. The models used for training our model are:

- Sequential Model
- VGG16
- VGG19
- ResNet50

We used **Adam optimizer** for updating the weights and learning rate as it combines the best properties of the AdaGrad and RMSProp **algorithms** to provide an optimization **algorithm** that can handle sparse gradients on noisy problems.

The **activation function** used is **Softmax** as our model is a multi-class classification problem.

Evolution of Models

SL No.	Year	Title Name/ Author Name	Model	Modality	Dataset	Results / Accuracy
1	2014	(Neverova et al., 2014)	CNN	2D, Depth	proposed dataset	82.0
2	2014	(Tosh & Szegedy, 2014)	DNN	2D, RGB	FLIC, LSP	96.0 , 78.0
3	2015	(Kang et al., 2015)	CNN	Depth	proposed dataset	99.0
4	2016	(Wei et al., 2016)	CNN	2D, RGB	MPII, LSP, FLIC	87.95, 4.32, 97.59
5	2017	(Wang et al., 2017)	CNN	2D, Depth	ChaLearn	55.57
6	2018	(Rao et al., 2018)	CNN	2D, Depth	Own dataset	92.88
7	2019	(Chen et al., 2019)	CNN	Dynamic, RGB	SHREC'17 Track Dataset	94.4
8	2020	(Wadhawan & Kumar, 2020)	CNN	Static, RGB	own dataset	99.72
9	2020	(Elboushaki et al., 2020)	CNN	Dynamic, RGB, Depth	SKIG	99.72
10	2021	Our model	CNN	Dynamic, Threshold	Own dataset	99.96

Results

The detailed mentioned in the above table are based on the Reference [1] given below. The total images considered for classification are approximately 30000. The VGG16 model has the highest accuracy rate compared with other models and loss is relatively low when compared to other models.

Model Name	Accuracy	Loss
Sequential Model	93%	0.48093
VGG16	99.96%	0.00114
VGG19	99.73%	0.00213
ResNet50	96.33%	0.47213

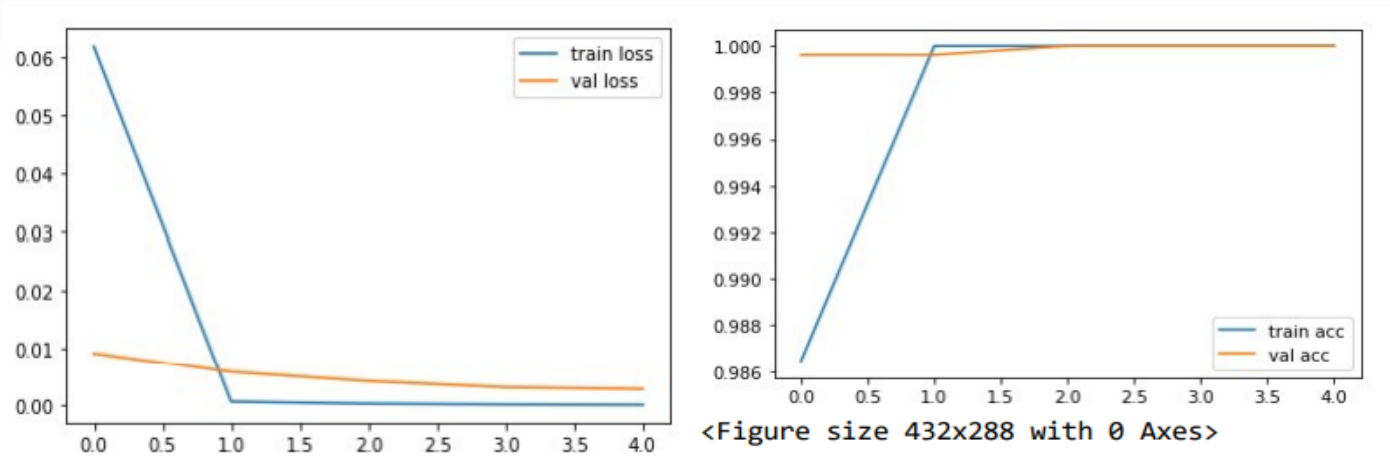


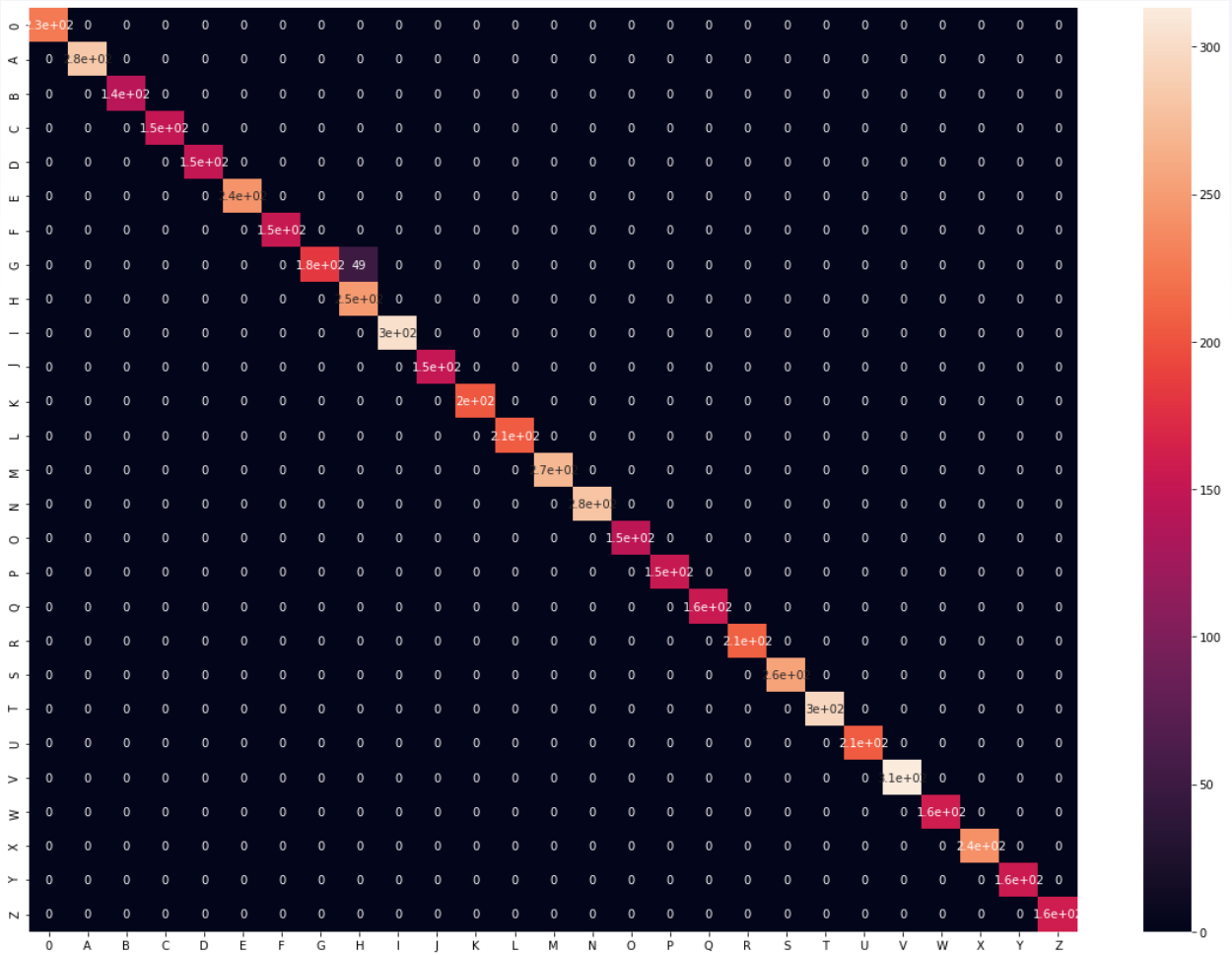
Fig.5: Model Loss and Model Accuracy curve

Classification Report and Confusion Matrix

The Classification report and Confusion matrix of the model VGG16 are given below for the detailed report of performance.

Classification Report

	precision	recall	f1-score	support
blank	1.00	1.00	1.00	226
A	1.00	1.00	1.00	284
B	1.00	1.00	1.00	142
C	1.00	1.00	1.00	150
D	1.00	1.00	1.00	150
E	1.00	1.00	1.00	244
F	1.00	1.00	1.00	154
G	1.00	0.79	0.88	232
H	0.84	1.00	0.91	249
I	1.00	1.00	1.00	304
J	1.00	1.00	1.00	151
K	1.00	1.00	1.00	203
L	1.00	1.00	1.00	206
M	1.00	1.00	1.00	271
N	1.00	1.00	1.00	280
O	1.00	1.00	1.00	146
P	1.00	1.00	1.00	153
Q	1.00	1.00	1.00	155
R	1.00	1.00	1.00	208
S	1.00	1.00	1.00	255
T	1.00	1.00	1.00	297
U	1.00	1.00	1.00	206
V	1.00	1.00	1.00	313
W	1.00	1.00	1.00	159
X	1.00	1.00	1.00	242
Y	1.00	1.00	1.00	156
Z	1.00	1.00	1.00	159
accuracy			0.99	5695
macro avg	0.99	0.99	0.99	5695
weighted avg	0.99	0.99	0.99	5695



Conclusion

The hand gesture recognition using state-of-the-art algorithms by transfer learning are implemented and achieved higher accuracy with less expensive and effortless methods. It is an efficient model which classify both for static and dynamic images. The predictions performed in this project are dynamic and used Adaptive Gaussian Thresholding with a deep neural network to get higher accuracy results. The recognition of hand gestures and conversion of text and speech is successfully implemented and the model can be used in the future thus by extending more functionalities for various predictions by feeding the model with required gestures. The future work can be accelerated by resolving the complications of the proposed model progressively and various techniques can be performed efficiently to overcome the limitations of the model.

References

1. Razieh Rastgoo, Kourosh Kiani, Sergio Escalera “Sign Language Recognition: A Deep Survey” , Electrical, and Computer Engineering Department, Semnan University, Semnan, 3513119111, Iran (2020, October).
2. Dhall, I., Vashisth, S., & Aggarwal, G. “Automated Hand Gesture Recognition using a Deep Convolutional Neural Network model”. 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (2020).
3. Cote-Allard, U., Fall, C. L., Campeau-Lecours, A., Gosselin, C., Laviolette, F., & Gosselin, B. “Transfer learning for serge hand gestures recognition using convolutional neural networks”. 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (2017).

Publication Details

- Presented paper in International conference, **9th ICCET - "International Conference On Contemporary Engineering And Technology" on 10th of April, 2021.**
- Paper Title:- "**Hand Gesture Recognition Using Transfer Learning Techniques**"
- Status:- **Accepted** by ICCET conference committee and paper would be published in reputed indexed journal soon.

Any Queries?

THANK YOU.