Paper 7.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 4- Issue 1,Paper 8, January 2021

HAND GESTURE RECOGNITION USING TRANSFER LEARNING TECHNIQUES

N Kumaran[1], M Sri Anurag[2] and M Sampath[3]
[1] Assistant Professor, Dept. of Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram, Tamil Nadu, India.
[2,3] Student, Dept. of Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram, Tamil Nadu, India.

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 paper focus on Automated Hand Gesture Recognition using transfer learning techniques.

Keywords — American Sign Language, hand gesture recognition, convolutional neural network, transfer learning.

I. INTRODUCTION

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, 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. Today, around one million people use American Sign Language (ASL) as their main way to communicate, according to Communication Service for the Deaf. This paper mainly focuses on finger spellings-based Hand Gestures of ASL.

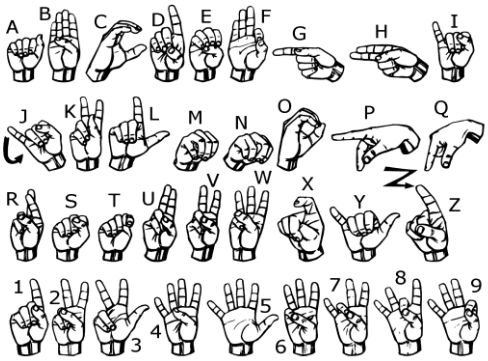

Fig.1: ASL Fingerspelling Alphabet

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1,Paper 8, January 2021

Fingerspelling is the representation of the letters of a writing system, and sometimes numeral systems, using only the hands. The best thing about using ASL is, it mostly uses single-handed gestures for the representation of finger spellings, which provide ease for the users. In this paper, we use 26 Alphabets of the English language for the representation of our model. We use Convolutional Neural Network (CNN) a class of deep learning which is mainly used for the classification of images and videos. CNN has brought a substantial change in the world of image classification as it provides a wide range of applications to train images using neural networks. It also provides **state of art algorithms like VGG16, VGG19, ResNet50, Inception V3**, and many more, which helps to transfer the parameters from pre-trained models to any proposed models.

II. LITERATURE SURVEY

In recent years there has been tremendous research done on hand gesture recognition. With the help of a literature survey done we realized the basic steps in hand gesture recognition are:

- Data acquisition
- Data preprocessing
- Feature extraction
- Gesture classification

2.1 Data acquisition:

The different approaches to acquire data about the hand gesture can be done in the following ways:

2.1.1 Use of sensory devices

It uses electromechanical devices to provide exact hand configuration and position. Different glove-based approaches can be used to extract information. But it is expensive and not user-friendly.

2.1.2 Vision-based approach

In vision-based methods, a computer camera is the input device for observing the information of hands or fingers. The main challenge of vision-based hand detection is to cope with the large variability of human hand's appearance due to a huge number of hand movements, to different skin-color possibilities as well as to the variations in viewpoints, scales, and speed of the camera capturing the scene.

Fig.2: Hand gesture Recognition System

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1,Paper 8, January 2021

2.2 Data preprocessing and Feature extraction for vision-based approach:

In [1] the approach for hand detection combines threshold-based color detection with background subtraction. We can use an Adaboost face detector to differentiate between faces and hands as both involve similar skin colors.

We can also extract the necessary image which is to be trained by applying a filter called Gaussian blur. The filter can be easily applied using open computer vision also known as OpenCV and is described in.

2.3 Gesture classification:

In [1] Hidden Markov Models (HMM) is used for the classification of the gestures. This model deals with the dynamic aspects of gestures.

In [2] Naïve Bayes Classifier is used which is an effective and fast method for static hand gesture recognition. It is based on classifying the different gestures according to geometric-based invariants which are obtained from image data after segmentation. Thus, unlike many other recognition methods, this method is not dependent on skin color.

The next step is the classification of gestures by using a K nearest neighbor algorithm aided with a distance weighting algorithm (KNNDW) to provide suitable data for a locally weighted Naïve Bayes‟ classifier.

According to a paper on "Human Hand Gesture Recognition Using a Convolution Neural Network" by Hsien-I Lin, Ming-Hsiang Hsu, and Wei-Kai Chen graduates of the Institute of Automation Technology National Taipei University of Technology Taipei, Taiwan, they construct a skin model to extract the hand out of an image and then apply binary threshold technique to the whole image.

After obtaining the threshold image they calibrate it about the principal axis to center the image about it. They input this image into a convolutional neural network model to train and predict the outputs. They have trained their model over 7 hand gestures and using their model they produce an accuracy of around 95% for those 7 gestures.

Based on research done we found that these alphabets frequently fail the prediction of the model, in this approach we try to overcome these limitations.

- For D: R and U
- For U : D and R
- For I: T, D, K, and I
- For S: M and N

The information provided in the Table.1 is based upon [4]. It contains the various techniques used by various authors and how the results are significantly increasing over time.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1,Paper 8, January 2021

Table.1: comparison of various model results overtime

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

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1,Paper 8, January 2021

III. METHODOLOGY AND ALGORITHM

Presently, our model acquires 99.96% of accuracy which will be beneficial for the model to be used for intended purposes. Also, our model is limited to good lightning conditions with a static background which would help our model to predict at its best. Research is being conducted on this field and some questions remain unsolved. Our model is consisting of three modules i.e., data generation, pre-processing, gesture classification.

3.1 Input Gesture

In [9] the gesture of the hand are detected using hand segmentation techniques which limits to different skin colors and with good lighting conditions. The combination of Adaptive Gaussian Thresholding with Otsu's Thresholding helps the image to predict the edges accurately and makes the image to be independent of skin color. But, it limits with a static background which is the only drawback in applying this technique. This paper aims to achieve better results with limited resources as that the Thresholding technique gives high results as compared with the hand segmentation. Fig.3 shows the threshold image which is considered as input for the model.

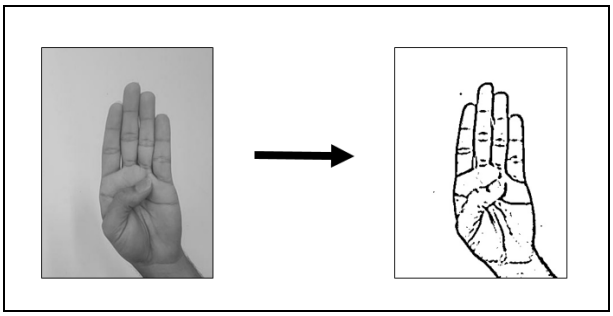

Fig.3: Adaptive Gaussian Threshold image

3.2 CNN Algorithm

Neural Networks are the most efficient way for replication of the human brain, they can gain as high accuracy as humans. These neural networks act as neurons and can learn at a very deep extinct. Convolutional Neural Networks are the most widely used neural class that is applied to images and videos. It is quite similar to Artificial Neural Networks using machine learning algorithms.

Convolutional Neural Network consists of a convolutional layer which is used to extract the various features from the input images, then it is followed by a pooling layer. The primary aim of this layer is to decrease the size of the convolved feature map to reduce the computational costs, and finally, a fully connected layer takes the high-level images from the previous layer's filtered output and converts them into a vector. Each layer of the previous matrix will be first converted into a single (flatten) dimensional vector then each vector is fully connected to the next layer that is connected through a weight matrix [7]. We have various hidden layers in between the convolutional layer and fully connected layers.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1,Paper 8, January 2021

Softmax is an activation function which is especially used in classification purpose. As we are up to a multi-class classification it helps to find a class of each digit and generate output. Adam optimizer is used for updating the model in response to the output of the loss function. Adam combines the advantages of two extensions of two stochastic gradient descent algorithms namely adaptive gradient algorithm (ADA GRAD) and root mean square propagation (RMSProp).

3.2.1 Sequential model

A Sequential model is appropriate for **a plain stack of layers** where each layer has **exactly one input tensor and one output tensor [8].** We create a Sequential model by passing a list of layer instances to the constructor, we created a model by adding up the convolutional layer, pooling layer, Dropout layer (which prevents the problem of overfitting), and dense layer which ended up getting a satisfying result of accuracy around (80-90) percentage.

3.3 Pre-trained models

3.3.1 VGG16

VGG16 is a convolution neural network (CNN) architecture with a pre-trained very deep neural network for large-scale image recognition. It consists of 13 convolutional layers and 3 fully connected layers with 138,357,544 parameters. It is a widely used neural network that trains and predicts images with 3 channel RGB inputs and outputs. As our input is of 1 channel threshold image, we converted it into 3 channel image and feed to VGG16 and got up with magnificent results of 99.96% of accuracy. The model is accurate in predicting 26 hand gestures of American Sign Language.

3.3.2 VGG19 & ResNet50

A similar operation is conducted using VGG19 & ResNet50 neural networks which share the similar architecture of VGG16 with extra added layers and a different set of parameters. After a lot of research, we found VGG16 performs better in recognition of hand gestures when compared with others.

IV. RESULTS

The experimental results of this paper show that the model proposed here extract features from the input images and classify various gestures of American Sign Language with greater accuracy of 99.96% with negligible loss of 0.00114. The model is providing a remarkable results on 26 finger spellings of ASL. It is much faster and more accurate when compared to other models.

Fig.4 illustrates the results of hand gestures of American Sign Language classified by pre-trained neural network VGG16. Fig.5 illustrate the accuracy and loss curve of our model.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1,Paper 8, January 2021

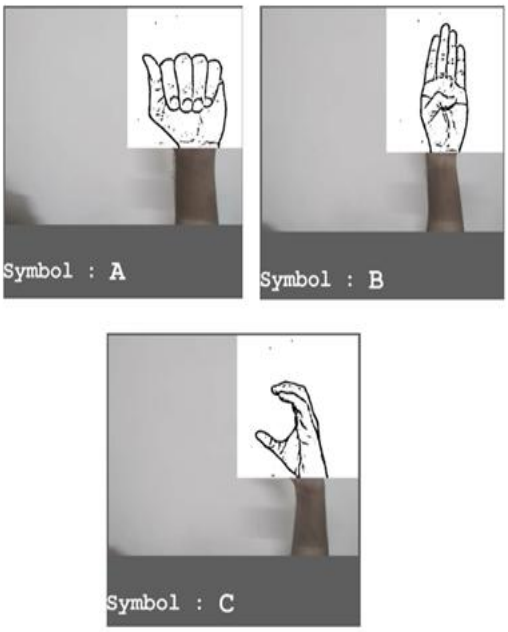

Fig.4: Our model predicting Hand Gesture of American Sign Language (A, B, C)

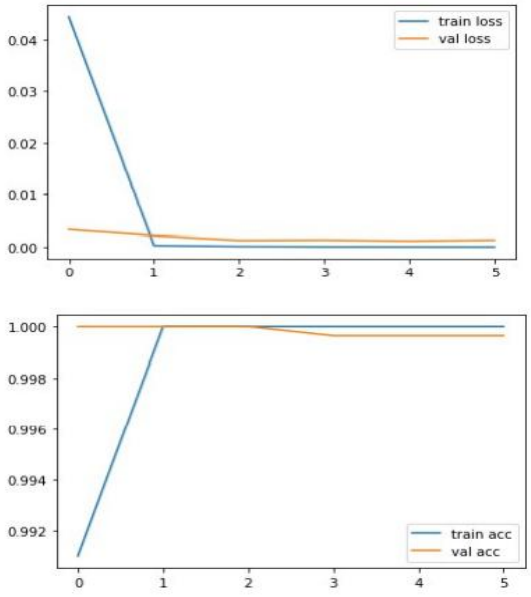

Fig.5: model accuracy curve and model loss curve

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 1,Paper 8, January 2021

Achieving a higher accuracy than existing models of 99.96% is quite challenging, this is possible by tuning Hyperparameters and with the help of data preprocessing techniques.

V. CONCLUSION AND FUTURE WORK

This paper focuses on the recognition of hand gestures using the state-of-the-art algorithm and achieve higher accuracy with less expensive and easy methods. It is an efficient model which can be used both for static and dynamic images. The predictions performed in this paper 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 by extending more functionalities for various predictions by feeding the model with required gestures and training to get more classes.

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] T. Yang, Y. Xu, and "A., Hidden Markov Model for Gesture Recognition", CMU-RI-TR-94 10, Robotics Institute, Carnegie Mellon Univ., Pittsburgh, PA (May 1994).
[2] Pujan Ziaie, Thomas M¨uller, Mary Ellen Foster, and Alois Knoll"A Na¨ive Bayes Munich, Dept. of Informatics VI, Robotics and Embedded Systems, Boltzmannstr. 3, DE-85748 Garching, Germany.
[3] Chen, X., Li, Y., Hu, R., Zhang, X., & Chen, X. (2020). Hand Gesture Recognition based on Surface Electromyography using Convolutional Neural Network with Transfer Learning Method (2020).
[4] Razieh Rastgoo, Kourosh Kiani, Sergio Escalera (2020, October). Sign Language Recognition: A Deep Survey, Electrical, and Computer Engineering Department, Semnan University, Semnan, 3513119111, Iran.
[5] Wadhawan, A., & Kumar, P. Sign Language Recognition Systems: A Decade Systematic Literature Review. Archives of Computational Methods in Engineering (2019).
[6] Murtaza, Z., Akmal, H., Afzal, W., Gelani, H. E., ul Abdin, Z., & Gulzar, M. H. Human-Computer Interaction Based on Gestural Cues Recognition/Sign Language to Text Conversion. In 2019 International Conference on Engineering and Emerging Technologies (ICEET) (pp. 1-6). IEEE (2019, February).
[7] Ren, Zhou, et al. "Robust part-based hand gesture recognition using Kinect sensor." IEEE transactions on multimedia 15.5, pp.1110-1120, (2013).
[8] Singh, J. P., Gupta, A., & Ankita. (2020). Scientific Exploration of Hand Gesture Recognition to Text. International Conference on Electronics and Sustainable Communication Systems (ICESC) (2020).
[9] 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).
[10] nuja P. Parameshwaran, Heta P. Desai, Rajshekhar Sunderraman , Michael Weeks, Georgia State University. Transfer Learning for Classifying Single Hand Gestures on Comprehensive Bharatanatyam Mudra Dataset (2019).