Sign Language Decoder using Convolutional Neural Networks

Malyala Sri Vikas[1], Pothunoori Shiva Kumar[2], Kolanupaka Sushank[3], K. Jeevan Reddy[4]

[1,2,3]Sreenidhi Institute of Science and Technology, Ghatkesar
[4]Associate Professor, Dept. of Electronics and Communication Engineering, Sreenidhi Institute of Science and Technology College, Telangana, India.

Abstract - *There exists an communication problem between the deaf community and the hearing majority. Sign language acts as a bridge between the speech and the hearing impaired community. It is difficult for most people who are not familiar with sign language and is not understood by everyone to communicate without an interpreter. In the era of multi modes for communication, sign language is, and continues to be, one of the most understudied areas. Automatic recognition of gestures using computer vision and machine learning is important for many real world applications such as sign language recognition. We hereby present the development and implementation of an American Sign Language (ASL) decoder based on a convolutional neural network. In this paper, we present a method for using deep convolutional neural networks to classify images of the letters in American Sign Language.*

***Key Words*: OpenCV, Machine Learning, Convolution Neural Networks, Deep Learning, Gesture Recognition, Sign language recognition.**

1. INTRODUCTION

Sign language recognition is vital for natural and convenient communication between deaf community and hearing majority. Currently, most communications between two communities highly depend on human-based translation services. However, this is often inconvenient and expensive as human expertise is involved. Therefore, automatic recognition aims to know the meaning of signs without the help from experts. Then it can be translated to sound or text as per the requirement of end users. Sign language recognition is still a challenging problem despite of many research efforts during the last few decades. One of the problems in sign language is how often fingerspelling is used. Fingerspelling is a method of spelling a word using only hand gestures. One of the explanations of why the fingerspelling alphabet plays such a vital role in sign language is that signers used it to spell out names of anything for which there is not a sign. Moreover, even same signs have significantly different appearances for different signers and different viewpoints. There are basically two sorts of approaches for hand gesture recognition: vision based approaches and data glove approaches. The main focus of this work is to make a vision based system to find Finger spelled letters of ASL. The reason for selecting a system supported vision relates to the very fact that it provides an easier and more intuitive way of communication between a person and a computer. In this paper, 26 categories for English Alphabets (a-z) are considered. In our research, we look at American Sign Language (ASL), which is used in the USA and in English-speaking Canada and has many different dialects.

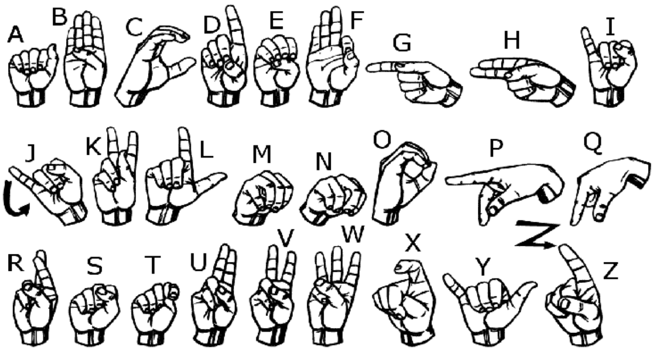

Fig -1: ASL (American Sign Language) 26 Hand Gestures.

In this paper, we have proposed a deep architecture with convolutional neural network for continuous sign language recognition. We have designed a staged optimization process for training our deep neural network architecture. We fully exploit the representation capability of CNN with tuning on vast amounts of gloss-level segments and effectively avoid overfitting with the deep architecture. We have also proposed a novel detection net for regularization on the consistency between sequential predictions and detection results. The effectiveness of our approach is demonstrated on a challenging benchmark, where we have achieved the performance comparable to the state-of-the art.

1.1 MOTIVATION

A Sign Language (SL) is the natural way of communication of deaf community. About 360 million people worldwide having hearing problem (about 5%). In recent survey of Indian Paediatrics 4 out of 1000 children born in India have hearing problem. The well-known sign languages are namely [7]:

- American Sign Language (ASL)
- Israeli Sign Language
- Indian Sign Language (ISL)
- Pakistani Sign language
- South Korean Sign Language
- Taiwan Sign Language
- Arabic Sign Language and so on

More than 1 million of adults and 0.5 million children in India make use of Indian Sign language[6].

1.2 RELATED WORK

Although gesture recognition only considers well specified hand gestures, some approaches are related to sign language recognition. Nagietal. [1] proposed a gesture recognition system for human-robot interaction using CNN. Van den Bergh et al. [2] proposed a hand gesture recognition system using Haar wavelets and database searching. The system extracts features using Haar wavelets and classifies input image by finding the nearest match in the database. Although both systems show good results, these methods consider only six gesture classes. Different sign languages are used in different countries or regions. There have been efforts towards sign language recognition systems other than ASLaswell. Pigouetal.[3] proposed an Italian sign language recognition system using CNNs. Although they reported 95.68% accuracy for 20 classes, they mentioned that users in test set can be in training set and/or validation set. Liwicki et al. [4] described a British Sign Language recognition system that understands fingerspelled words from video. The system first recognized letters using Histogram of Gradients (HOG) descriptors. Then that recognized words using Hidden Markov Models (HMM). That system is different from recognizing a single fingerspelling. The dataset in use corresponded to a single signer.

Convolutional Neural Networks have been extremely successful in image recognition and classification problems, and have been successfully implemented for human gesture recognition in recent years. In particular, there has been work done in the realm of sign language recognition using deep CNNs, with input-recognition that is sensitive to more than just pixels of the images. With the use of cameras that sense depth and contour, the process is made much easier via developing characteristic depth and motion profiles for each sign language gesture[5].

2. BLOCK DIAGRAM

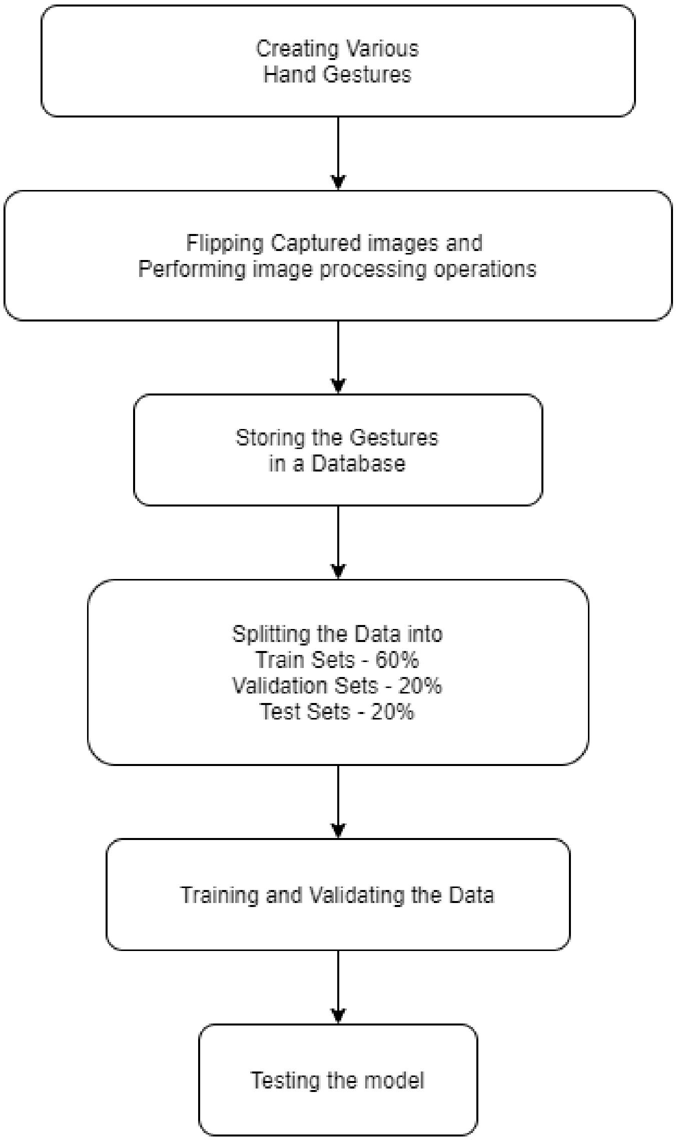

Fig -2 Block Diagram of the Gesture recognition system

3. METHODOLOGY

3.1 DATA

We have collected 31,200, Images of 26 alphabets from 3 people for each alphabet except J and Z which require temporal information for classification, with a camera of resolution of 720p (1280×720). Since our dataset was not captured in a controlled environment, it was prone to differences in light, colour of the skin, background colour and other differences in the environment. We also flipped the images vertically as we can input the algorithm much more efficiently there increasing the datasets to 62,400. We then stored the data into a database using SQLite3 i.e concepts of DBMS.

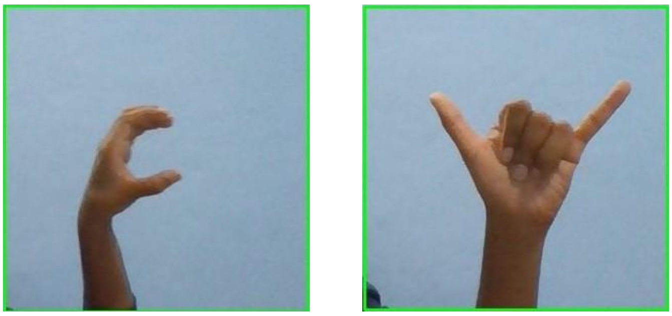

Fig -3 Gesture for Alphabets –C & Y.

3.2 DATA PRE-PROCESSING

After generating our own dataset, we then differentiate the background from the original data sets using thresholding from OpenCV. Once the thresholding is done we apply various filters to the images such as gaussian blurring to remove noise from the images and store them in the database. We then split the data into Training data sets which contain 52000 images (80%), Validation data sets which contain 5200 images (20%), Test data sets which contain 5200 images (20%).

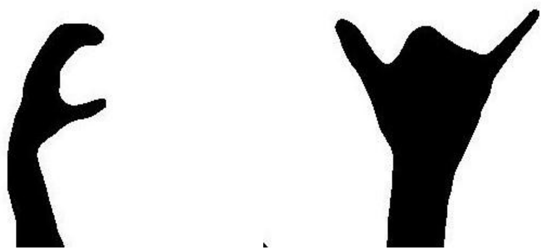

Fig -4 Processed Gesture for Alphabets –C & Y.

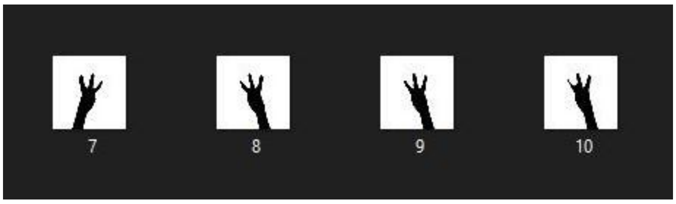

Fig -5 Images in the database.

Fig -6 Processed gestures for every alphabet.

3.3 TRAINING

During training, dropout [9] and data augmentation are used as main approaches to reduce overfitting. We have used a sequential keras model as our neural network model which contains one or more convolutional layers and followed by one or more Pooling layers as in a standard multilayer neural network. The minimum batch size of 400 and Epoches of 25 was choosen. The learning rate is initialized at 0.03 with a 4.5% decrease after each epoch. The weights were randomly initialized. Training is performed on one machine with a hexa-core processor (Intel Core i5-3930K), 8GB SDRAM and a NVIDIA GeForce GTX 680 GPU with 2GB of memory . The models are implemented using the Python libraries Theano [8], and Keras. We could achieve a training accuracy of 85% without over fitting the model.

Convolutional layer: Core building block of a CNN. Layer's parameters have of a set of learnable filters (or kernels). Filter is convolved across the length and breadth of the input volume. Computes the dot product between the inputs of the filter

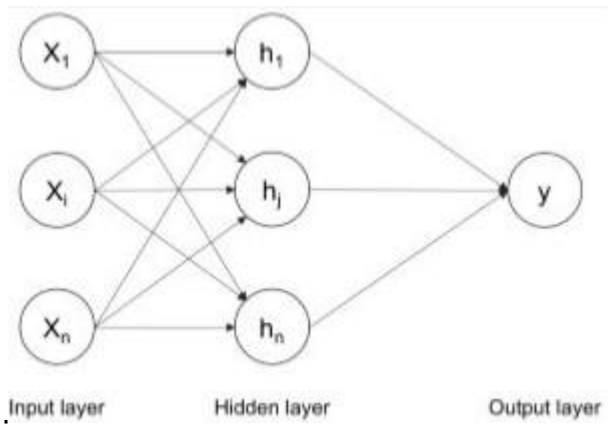

Fig -7 Convolutional Layer

Pooling Layer: This Layer Reduces the spatial size of the representation to decrease the number of parameters. Independently acts on every depth slice of the input. Most common type is a pooling layer with filters of size 2x2 applied with a stride of 2 down samples every depth slice in the input by 2 along with both the width and the height, discarding 75% of the activations spatially, using the MAX operation.

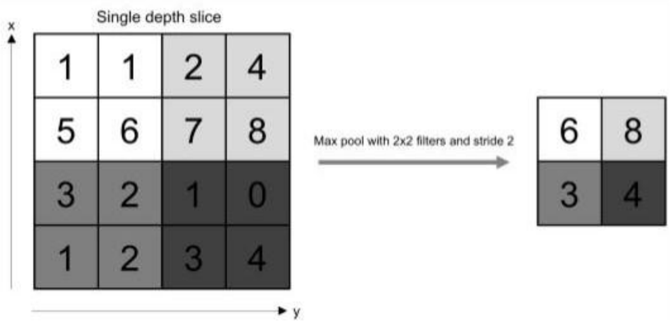

Fig -8 Pooling Layer.

ReLU layer: ReLU stands for Rectified Linear Units which increases the nonlinear properties.

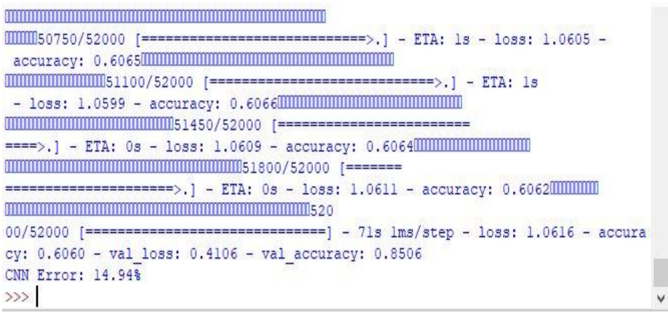

Fig -9 Accuracy of the Trained Model

4. RESULTS

The system was tested on all the 26 different alphabets for each sign and every sign is decoded accurately. The accuracy was checked as per correctness of every Alphabet. The maximum accuracy is 97% for all the alphabets. This implies that the system works efficiently for most of the alphabetic character recognitions. Since this model does not use the temporal information classification the gestures j and z were not decoded.

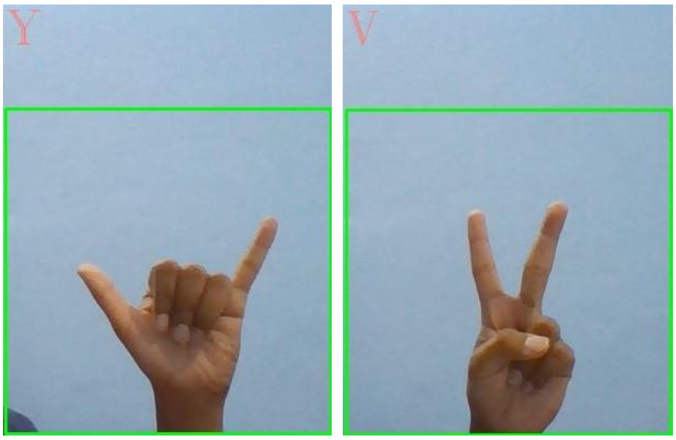

Fig -10 Output when gesture is shown.

5. CONCLUSIONS

This paper deals with the application of Convolution Neural Network for recognizing the hand gestures. We described a deep learning based approach for a classification of American Sign Language. This work shows that convolutional neural networks can be used to accurately recognize different signs of a sign language. One of the major applications of hand gesture recognition is to identify the sign language which is a important tool of communication for physically impaired, deaf and dumb people. This tool will help to bridge the gap between normal and deaf/dumb people. Our method shows that this model has enough potential in solving the problem using a simple web camera, if sufficient training data is provided.

From the results obtained above we can conclude that CNN provides a impeccable accuracy in identifying the sign language gestures. We have created an idea of translating the static image of sign language into a text format. The static image includes alphabet used in both training, validating and testing of data. Features representation will be learned by a technique known as convolutional neural networks. This work can be further extended to building a real time application which can identify the sign language gestures including words and sentences to recognize instead of just characters.

ACKNOWLEDGEMENT

We express our earnest gratitude to our Guide, **K. JEEVAN REDDY**, for his constant support, encouragement, guidance and motivating us in our endeavor and helping us to realize our full potential. We are grateful for his cooperation and valuable suggestions.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

REFERENCES

[1] J. Nagi, F. Ducatelle, G. Di Caro, D. Ciresan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. Gambardella. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on, pages 342–347, Nov 2011.

[2] M. Van den Bergh and L. Van Gool. Combining rgb and tof cameras for real-time 3d hand gesture interaction. In Applications of Computer Vision (WACV), 2011 IEEE Workshop on, pages 66–72, Jan 2011.

[3] L.Pigou, S.Dieleman, P.-J.Kindermans, and B.Schrauwen. Sign language recognition using convolutional neural networks. In Computer Vision - ECCV 2014 Workshops, pages 572–578, 2015.

[4] S. Liwicki and M. Everingham. Automatic recognition of finger spelled words in british sign language. InComputerVisionandPatternRecognitionWorkshops,2009.CVPRWorkshops 2009. IEEE Computer Society Conference on, pages 50–57, June 2009.

[5] Agarwal, Anant & Thakur, Manish. Sign Language Recognition using Microsoft Kinect. In IEEE International Conference on Contemporary Computing, 2013.

[6] N. Purva, K. Vaishali, "Indian Sign language Recognition: A Review", IEEE proceedings on International Conference on Electronics and Communication Systems, pp. 452-456, 2014.

[7] F. Pravin, D. Rajiv, ""HASTA MUDRA" An Interpretation of Indian Sign Hand Gestures", 3rd International conference on Electronics Computer technology, vol. 2, pp.377-380, 2011.

[8] Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., Bengio, Y.: Theano: a CPU and GPU math expression compiler. In: Proceedings of the Python for Scientific Computing Conference (SciPy) (Jun 2010), oral Presentation.

[9] Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012).

BIOGRAPHIES

Malyala Sri Vikas,
Studying at Sreenidhi institute of science and technology, Dept. of Electronics and Communication Engineering.

Pothunoori Shiva Kumar,
Studying at Sreenidhi institute of science and technology, Dept. of Electronics and Communication Engineering.

Kolanupaka Sushank,
Studying at Sreenidhi institute of science and technology, Dept. of Electronics and Communication Engineering

K Jeevan Reddy,
Associate professor at Sreenidhi institute of science and technology, Dept. of Electronics and Communication Engineering