# American Sign Language Detection Using Convolutional Neural Network

Deep Dand,

(Assisted by Shruti Kulkarni)

Marist College

***Abstract***

American sign language is method of communicating among the speech and hearing impaired community.It is difficult for most people to communicate who are not familiar with sign language or without an interpreter. Innovations in sign language recognition try to tear down this communication barrier. Our contribution is to propose a recognition system using convolutional neural networks (CNNs) and GPU acceleration to identify hand-sign gestures. CNNs are able to learn and identify the gestures from statics images. We are working with the 26 letters of English alphabets and images captured in grayscale with different architectures of the CNN to achieve the highest accuracy.

Keywords - cnn, gpu, slrs, asl

## ***Introduction***

Sign language is one of the alternative and expressive method used for communicating between hearing impaired people where information is conveyed through hand gestures. These hand-gestures is an information base to learn and evolve for any Sign-Language-Recognition-System (SLRS). We propose, in building such system with two main steps-The first step is to extract features from the static images of hand gestures. This will result in a representation consisting of one or more feature vectors. This representation will aid the system to distinguish between the possible classes of actions in this case 31 classes. The second step is the classification of the action(gestures from one of 31 classes). A classifier will use these representations to differentiate between the various actions (or signs). In our work, the classification and feature extraction is automated by using convolutional neural networks (CNNs).

### *Related Work*

ASL detection is a topic which is active in research from last two decades. With current research about past experiments, I have found that Nachanau. M[1] has implemented the Alphabet recognition of ASL using the SIFT algorithm. The approach is to take 26 English alphabets. The database is of the color images and has no background to it after the image pre-processing. After pre-processing, the signature is created from the extracting features and finding key points. Using feature vector composition and SIFT algorithm, they achieved 100% accuracy on test samples of size 16*16 image data.

Brandon Garcia and Sigberto Alarcon Viesca[2] have utilized pre-trained GoogLeNet architecture trained on ILSVRC2012 dataset. Using SVM, Caffe framework they transferred the learning. For pre-processing the data was resized to 256*256 and performed padding with black pixels to eliminate the difference between two classes in existing dataset with respect to ILSVRC2012 dataset. The dataset was divided in subnets a-y, a-k and a-e and achieved top-1 validation accuracy of 0.72, 0.74 and 0.97 respectively.

### *Methodology*

[3]Our current architecture consists of one input layer, 2 convolutional layer- for extracting hand features and we use max-pooling method to simplify the image, one fully connected layer and output layer.

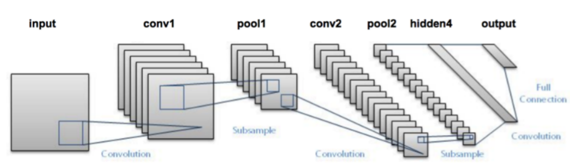


Fig.1 [5] The architecture of 2-layer CNN.

The 2D convolution is defined as -

$$I' = \sum_{u,v} I(x-u, y-v)\, g(u,v)$$

Weights and biases

$$\sigma\left(b + \sum_{l=0}^{4} \sum_{m=0}^{4} w_{l,m}\, a_{j+l,k+m}\right)$$

***Architecture***

- Number of Layers - 2
- Layer 1
  - 32 filters
  - Relu activation
  - Maxpool filter size- 1
- Layer 2
  - 64 Filters
  - Relu activation
  - Maxpool filter size- 2
- Fully Connected Layer -
  - 512-1024 units.
  - Softmax activation
- Dropout layer - configured to drop results less than 0.8
- Number of Classes - 31
- IMG_SIZE = 256x256
- Learning rate - 1e3

### *Experiments*

Currently we are experimenting model with different value of number of convolution layers,ReLus,dropout, Learning rate, filter and kernel size, number of epochs to optimize the error-loss and gain highest accuracy with data set of 5 subjects.

We started off with minimum values for different parameters:

| Filter | | Maxpool filter | | Units in Fully Connected layer | Epochs | Test Accuracy | Val Accuracy |
|---|---|---|---|---|---|---|---|
| L1 | L2 | L1 | L2 | | | | |
| 3 | 3 | 1 | 2 | 512 | 5 | 0.04 | 0.03 |
| 5 | 5 | 1 | 1 | 512 | 3 | NA | NA |
| 5 | 5 | 1 | 1 | 756 | 3 | NA | NA |
| 5 | 5 | 1 | 2 | 512 | 3 | 0.03 | 0.03 |
| 5 | 5 | 1 | 2 | 512 | 5 | 0.03 | 0.03 |
| 3 | 3 | 3 | 3 | 1024 | 5 | 0.03 | 0.03 |
| 3 | 5 | 3 | 5 | 1024 | 5 | 0.037 | 0.032 |
| 5 | 3 | 5 | 3 | 1024 | 5 | 0.098 | 0.095 |
| 5 | 5 | 5 | 5 | 1024 | 5 | 0.9 | 0.9 |

### *Analysis*

We currently are working on complete image sizes to get a workable predictive model. The existing network works with really small images and so it with accuracy of 98.95% with just 5 epochs.

The program under execution currently is having image size of 256 and epochs 50 with the learning rate 1e-4 and is in 6th epoch. Once this finishes, we are expanding the architecture with 4 layers of feature extraction and vary number of epochs to see with what architecture we achieve the best accuracy. Once we have an architecture decided, we will use the complete dataset of 31000 images and divide 70-30 ratio for train and test.

### *Future Enhancements:*

- Since the images we are dealing with are 256x256, adding more convnet layers will improve feature extraction and result in the better accuracy of the model.
- Also, once the experiments complete with existing config, the next config will have more epochs with architecture that gives best result with 5 epochs.

### *Conclusion*

We desire to build an SLRS using CNN that can used to accurately recognize different hand gestures of different signs of known and unknown subjects with high performance. This generalization capacity of CNNs can contribute to broader research in field of ASL recognition.

*References*

1) M, N. (2013, January). ALPHABET RECOGNITION OF AMERICAN SIGN LANGUAGE: A HAND GESTURE RECOGNITION APPROACH USING SIFT ALGORITHM. Retrieved November 13, 2017, from http://aircconline.com/ijaia/V4N1/4113ijaia08.pdf
2) Garcia, B., & Viesca, S. A. (n.d.). Real-time American Sign Language Recognition with Convolutional Neural Networks. Retrieved November 13, 2017, from http://cs231n.stanford.edu/reports/2016/pdfs/214_Report.pdf
3) Deshpande, A. (n.d.). A Beginner's Guide To Understanding Convolutional Neural Networks. Retrieved November 13, 2017, from https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/
4) Priyadharsini, N., & Rajeswari, N. (2017, June). Sign Language Recognition Using Convolutional Neural Networks. Retrieved November 13, 2017, from http://www.ijritcc.org/download/browse/Volume_5_Issues/June_17_Volume_5_Issue_6/1497975929_20-06-2017.pdf
5) Pigou, L., Dieleman, S., Kindermans, P., & Schrauwen, B. (n.d.). Sign Language Recognition using Convolutional Neural Networks. Retrieved November 13, 2017, from https://pdfs.semanticscholar.org/ee72/744bea3c77a8f490d766d87517b1a450d44b.pdf