Mnist sign English language (ISIC) Image Classification with Convolutional Neural Networks

# By: Arad Zekler, 3005600579; Ido Kahana,316606680;

# Submitted in 15/01/2019

# DEEP LEARNING

**Abstract**

Wiki: "Sign languages (also known as signed languages) are [languages](https://en.wikipedia.org/wiki/Language) that use the visual-manual modality to convey meaning. Sign languages are expressed through manual articulations in combination with [non-manual elements](https://en.wikipedia.org/wiki/Sign_language#Non-manual_elements). Sign languages are full-fledged natural languages with their own grammar and lexicon. Sign languages are not universal and they are not [mutually intelligible](https://en.wikipedia.org/wiki/Mutual_intelligibility) with each other. although there are also striking similarities among sign languages."[4]

English sign symbols of characters consist of 25 static simple hand gesture for characters A-Z excluding the character J, and a simple motion hand gesture for the character J. Therefore we assuming that the English sign symbols classification problem is similar to other problem such as (Digit classification etc..) and the classification problem can be solve using existing well known method for image classification, in this article we will try to automated the classification of a sign images using a different standard variants of Convolutional Neural network and try to search for the best method, The best result in our experiments was achieved by Smaller VGGNET ARCHICTURE after 100 epochs training with result of 0.9936 of accuracy of test and 0.9995 accuracy of train

**Introduction**

In the following article we are going to test the data using 3 different Convolutional neural network (CNN) based models for Image classification problem with the default test and train data that was provided in the data source Kaggle data source.

1. The first model is a basic Model described in MNIST TensorFlow tutorial and are used to train the classic MNIST problem it contains 4 simple layers excluding the input.[2]
2. The second model is basic model that extend the well-known Convolutional neural network (CNN) model with 8 layers.
3. The third model is smaller VGG model (CNN) Based model with 20 layers[3]

|  |  |  |
| --- | --- | --- |
| Figure Basic Model | Figure Basic CNN | Figure VGN Model |

**Related Work**

There is no official related work to this problem, how ever this problem is very similar to classic MNIST Digit's classification problem. MNIST digit classification problem using the models from this article achieve the following results: the first model it achieve around 0.98 accuracy, using the basic CBB the best result was around 0.99 and in the VGG model 0.996 accuracy.

**Background Information**

A **Convolutional Neural Network (CNN)** is a Deep Learning algorithm which can take in an input image and assign importance (weights and biases) to different aspects in the image and later to be able to differentiate one images from the other. This ability can be used to classify the different parts in the image and output a diagnostic (-an accurate classification) to a English character of a specific hand gesture.

VGG is a Convolutional Neural Network architecture, It was proposed by [Karen Simonyan](http://www.robots.ox.ac.uk/~karen/) and [Andrew Zisserman](https://en.wikipedia.org/wiki/Andrew_Zisserman)of [Oxford Robotics Institute](https://en.wikipedia.org/wiki/Oxford_Robotics_Institute) in the the year 2014. It was submitted to Large Scale Visual Recognition Challenge 2014 (ILSVRC2014) and The model achieves 92.7% top-5 test accuracy in ImageNet. [ImageNet](https://en.wikipedia.org/wiki/ImageNet) is one the on the largest data-set available. It has 14 million hand-annotated images of what is in the picture.

**Project Description**

This project will utilize a number of different CNN models in trying to optimize the predicting accuracy of the type of character the gesture represent of a random hand gesture.

**The Database**

The dataset format is patterned to match closely with the classic MNIST. Each training and test case represents a label (0-25) as a one-to-one map for each alphabetic letter A-Z (and no cases for 9=J or 25=Z because of gesture motions). The training data (27,455 cases) and test data (7172 cases) are approximately half the size of the standard MNIST but otherwise similar with a header row of label, pixel1,pixel2....pixel784 which represent a single 28x28 pixel image with grayscale values between 0-255. The original hand gesture [image data](https://github.com/mon95/Sign-Language-and-Static-gesture-recognition-using-sklearn) represented multiple users repeating the gesture against different backgrounds. The Sign Language MNIST data came from greatly extending the small number (1704) of the color images included as not cropped around the hand region of interest. To create new data, an image pipeline was used based on ImageMagick and included cropping to hands-only, gray-scaling, resizing, and then creating at least 50+ variations to enlarge the quantity. The modification and expansion strategy was filters ('Mitchell', 'Robidoux', 'Catrom', 'Spline', 'Hermite'), along with 5% random pixelation, +/- 15% brightness/contrast, and finally 3 degrees rotation. Because of the tiny size of the images, these modifications effectively alter the resolution and class separation in interesting, controllable ways.[1]

**Experiment Results**

|  |  |  |
| --- | --- | --- |
| **Graph** | **Examples** | **Experiment** |
|  |  | The first experiment used a simple Basic architecture with an output sigmoid layer, and 1000 epoch trains, reached at is maximum a test accuracy of 0.66. and it look like the model is not fit. A bad result there is no points to continue with the experiments using this model. |
|  |  | The second experiment using the basic CNN Network and 10 Epochs showed a stunning improvement and reaching about 0.91 accuracy, it look like a potential good model |
|  |  | The Third experiment using the basic CNN Network and 100 Epochs showed a small improvement and reached to 0.92 accuracy at his max, I believe there is not point to continue with this model |
|  |  | The 4 experiment using a VGN CNN and 30 epochs showed a significance improvement from first EPOCH an reached an accuracy of about 0.98 at his max. |
|  |  | The 5 experiment using a VGN CNN and 100 epochs, showed a great progress and at his max the test accuracy was 0.9936 and the train was 0.9995 |

**Conclusions**

From utilizing different CNN architectures, it was demonstrated that this problem is actually similar to MINST problem and can be solved using the models that was used by MNIST digit classification problems, in out experiments we reached a peak of 0.9936, and show that it is possible using existing Techinic to automated and autodetect English sign hand gestures.

**References**

[1] **Sign MNIST DATA SET**   
<https://www.kaggle.com/datamunge/sign-language-mnist>

[2] **Tensor flow MNIST Basic classification** <https://www.tensorflow.org/tutorials/keras/classification>

[3] **Keras and Convolutional Neural Networks (CNNs),** Adrian Rosebrock, April 16, 2018.

[4] **Sign language**   
<https://en.wikipedia.org/wiki/Sign_language>.