Machine Learning Engineer Nanodegree

Capstone Proposal

Trong Canh NGUYEN

May 2017

Proposal

In this project we investigate the classification of the American Sign Language Letters from some image datasets using Deep Neural Network.

Domain Background

American Sign Language (ASL) is the language used by people with hearing and/or speaking disability. It used mainly hand shape, motion and sometimes facial expression as well.

In this project we will classify only a small subset of this language which is the Alphabet Letters (a-z) by using Deep Neural Network.

There are some reasons where I will focus only on this small subset:

- It is the good start to investigate the usability of deep learning for this ASL recognition problem. Future extension can be recognition of more general words.
- Some datasets are publically available.
- Similar researches on the same datasets were undertaken (with or without the same technique of deep learning) like in [1]. So we can compare our approach.

Problem Statement

In ASL each alphabet letter corresponds to a specific hand shape.

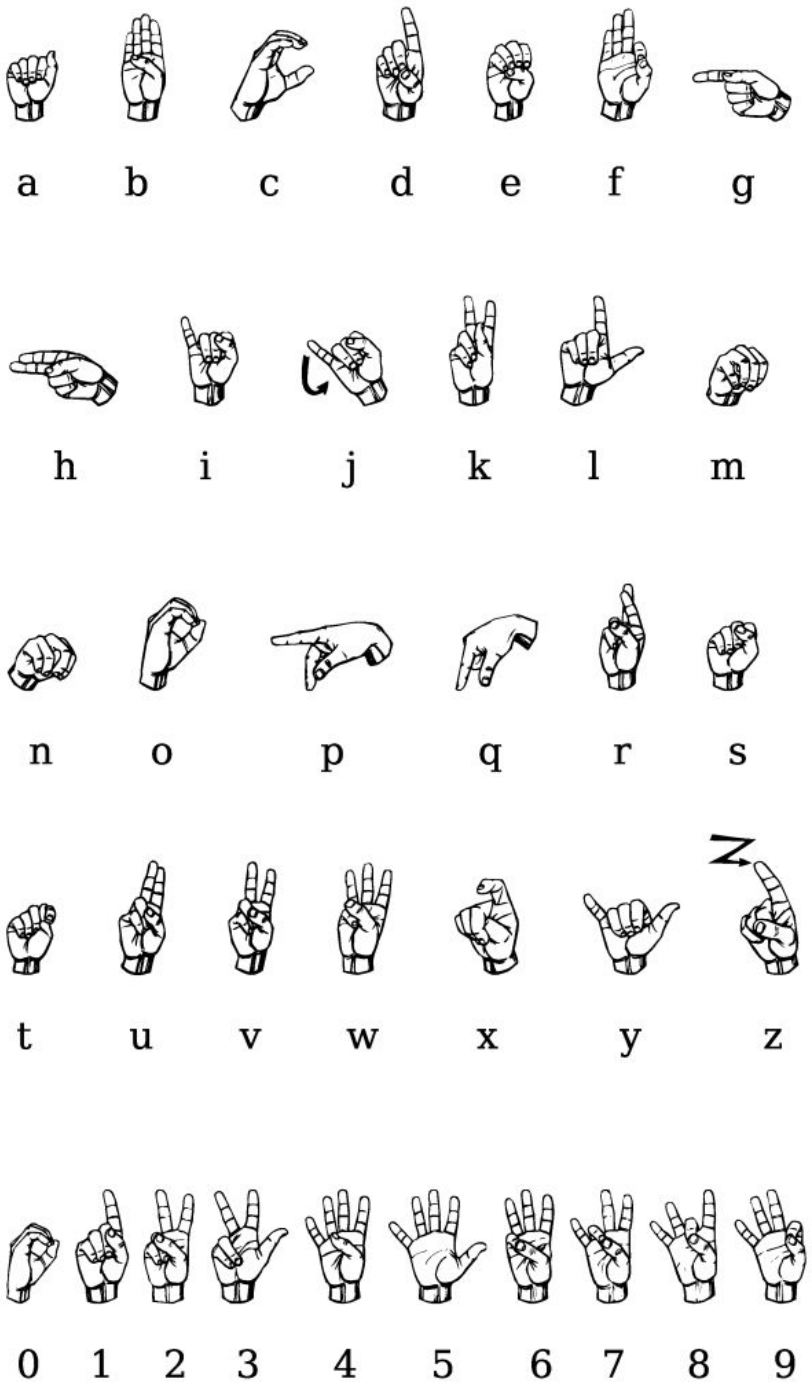

Image taken from Wikipedia: https://en.wikipedia.org/wiki/American_manual_alphabet

Most of them are static, except for J and Z which requires motion:

- J is I with a twist of the wrist, so that the little finger traces the curve of the printed form of the letter
- Z is an index finger moved back and forth, so that the finger traces the zig-zag shape of the letter Z

From images captured from a standard camera, each image is a hand shape for a letter, and our task is to recognize the letter from the image.

Since we use only static images, letters J and Z may be excluded from this study.

Some researches (like in [1]) use depth images captured by devices like Microsoft Kinect for example, which have more information. In this project we just use normal image. Depth images could be subject to future studies.

The problem is a **image classification problem**, similar to the digit recognition problem using MNIST dataset for example. Our objective is to know if the same deep learning method can be efficient with this dataset of hands shapes.

Datasets and Inputs

There are two datasets that are free and publicly available:

1. The first one is described in the reference [1] and the website below

http://empslocal.ex.ac.uk/people/staff/np331/index.php?section=FingerSpellingDataset

This dataset comprises 24 static signs (excluding letters j and z because they involve motion). This was captured in 5 different sessions, with similar lighting and background.

The dataset contains over 65,000 images, thus about 2700 by letter.

Here are images of the same letter **a**, from 5 sessions:

2. The second dataset [2] is provided by the Massey University and can be downloaded from

http://www.massey.ac.nz/~albarcza/gesture_dataset2012.html

This dataset contains closed-up images that are cropped such that the hand shape fits the border of the image. The background is changed to back. The images are taken from 5 different peoples. There are about 2500 images.

The images in our dataset are not of the same size, so some pre-work like resizing should be done.

Solution Statement

We will use deep neural network and convolutional neural network to classify the images. The use of deep neural network and convolutional neural network has encountered very successful results, even for small problem and dataset like the digit classification with MNIST dataset and thus worth to attempt for our problem as well.

Benchmark Model

For this image classification problem we will compare our results with the results from the paper [1] which use the same dataset. They provide a confusion matrix and thus we can compare our results quantitatively.

Evaluation Metrics

The evaluation metric will be the confusion matrix we we can have information on:

- The precision and recall of detecting a letter
- Letters that are confused with each other, which may help identify problems and improve our model.

Project Design

1.First we will do some analysis and standardization of our dataset since the images have different size.

2. We will try with some standard deep neural network to have an idea on the performances and how deep neural network is adapted to this problem (dataset)

3. We then use Convolutional Neural Network to see if it can improve the results.

4. Further optimization (data augmentation for example by adding noise to the ground, image rotations...)

References:

[1] Spelling It Out: Real-Time ASL Fingerspelling Recognition In *Proceedings of the 1st IEEE Workshop on Consumer Depth Cameras for Computer Vision*, jointly with ICCV'2011: Pugeault, N., and Bowden, R. (2011).

[2] New 2D Static Hand Gesture Colour Image Dataset for ASL Gestures:
A.L.C. Barczak, N.H. Reyes, M. Abastillas, A. Piccio and T. Susnjak
IIMS, Massey University, Auckland, New Zealand