American Sign Language (ASL) Image Classification Using CNNs

FNU Ujjwal PF10610@UMBC.EDU
Data Science Department, University of Maryland, Baltimore County, 1000 Hilltop Circle, Sherman Hall East 4th Floor, Baltimore, MD 21250 USA

Abstract

The sign language is an important tool of communication for hard hearing community, and it varies from region to region. American Sign Language or ASL is prevalent for hearing or speech impaired people in North America. ASL image classification is finding its use in many mobile based applications that can help integrate deaf community better with the society. Convolutional Neural Networks or CNNs are being widely used to address this problem. In this paper, I aim to experiment with different CNN models and observe the effects of varying kernel size on CNN's capability to perform ASL image classification. It will be a multi-class single label classification problem in which the CNNs will classify a particular image into one of 29 classes. I also aim to find how 'Swish' and 'Mish' activation function compares to 'ReLU' activation function in CNN classification models.

1. Introduction

"Sign language enables the smooth communication in the community of people with speaking and hearing difficulty (deaf and dumb)" (V. & R., 2020). But, very few people among the society learn this language as it requires memorizing a lot of hand postures. This creates a communication gap between the general public and the hearing and speech impaired community.

Therefore, many researchers have come out with an idea of image classification apps that may help the general public to understand what a person using the sign language is trying to say. CNNs has been proven as an effective algorithm for image classification tasks. "CNNs automate the process of feature extraction by learning the high-level abstractions in images and capture the most discriminative feature values using hierarchical architecture" (V. & R., 2020). Therefore, finding the most optimal CNN model to perform image classification task on ASL dataset will be very helpful in boosting the performance of such apps.

I will be experimenting with CNN models by varying the kernel size and activation functions to find an optimal CNN model.

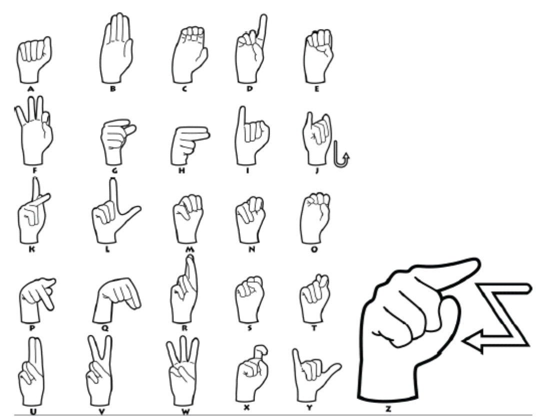

Figure 1. American Sign Language hand gestures for the alphabets (Daroya, Peralta, & Naval, Jr.)

2. Dataset to be used

I will be using 'ASL Alphabet – Image data set for alphabets in the American Sign Language' dataset available on Kaggle. It is a 1 GB dataset consisting of 87,000 images of size 200 x 200 pixels. There are a total of 29 classes and distribution of these classes is given as follows:

Table 1. Classification labels.

LABELS	INSTANCES
A-Z (26 ALPHABETS)	78,000 (3000 instances each)
SPACE	3000
DELETE	3000
NOTHING	3000

The link to the dataset is as follows:

https://www.kaggle.com/grassknoted/asl-alphabet

3. Project Idea

First, I aim to observe the effect of different kernel size on the performance of CNN model, keeping its architecture fixed. This means that I will only vary kernel size (like 3x3, 5x5, 7x7 etc.) and not any other aspect of the model (like activation functions or number of hidden layers).

Second, I plan to compare the performance of the CNN models when different activation functions are used. I am interested in exploring the potential of ‘Swish’ and ‘Mish’ as an alternate to ‘ReLU’. I intend to do this to understand the significance of dead neuron problem posed by ReLU.

Third, I will be utilizing some of the data augmentation techniques like horizontal shifting, vertical shifting, horizontal flip, vertical flip, random brightness augmentation etc. I will be doing this to observe the effects of data augmentation on performance of the CNN models.

4. Software

I will be using Python programming language to code, and I will be deploying my codes in Jupyter Notebook or Google Colab.

5. Plan of Work

First week will be dedicated towards pre-processing the data and preparing 2 datasets (1 non-augmented dataset and 1 augmented dataset). Second and third week will be dedicated to preparing the different CNN models for experimentation and subsequently training and testing them on the prepared datasets. Fourth and fifth week will be focused on evaluating the models and writing the final project report.

6. Research Papers

Number of research has been conducted on Image classification tasks using CNNs. Initially I will be going through the following research papers and materials to start my project:

Daroya, Rangel & Peralta, Daryl & Naval, Prospero. (2018). Alphabet Sign Language Image Classification Using Deep Learning. 0646-0650.10.1109 /TENCON.2018.8650241.

Adithya V., Rajesh R., A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition, Procedia Computer Science, Vol. 171, 2020, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.04.255.

Anamika Srivastava, Vikrant Malik (2020) Review on Sign Language Detection Using Machine Learning. Journal of Critical Reviews, 7 (10), 1190-1194. doi:10.31838/jcr.07.10.234

Fierro, Atoany & Perez-Daniel, Karina. (2020). Siamese Convolutional Neural Network for ASL Alphabet Recognition. Computación y Sistemas. 24. 10.13053/cys-24-3-3481.

Rao, G.A., Syamala, K., Kishore, P.V., & Sastry, A.S. (2018). Deep convolutional neural networks for sign language recognition. 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), 194-197.

Bheda, Vivek & Radpour, Dianna. (2017). Using Deep Convolutional Networks for Gesture Recognition in American Sign Language.

Grandhi, Chandhini and Sean Liu. “American Sign Language Recognition using Deep Learning.”

Garcia, Brandon. “Real-time American Sign Language Recognition with Convolutional Neural Networks.”

Pigou, L., Dieleman, S., Kindermans, P., & Schrauwen, B. (2014). Sign Language Recognition Using Convolutional Neural Networks. ECCV Workshops.

References

Daroya, Rangel & Peralta, Daryl & Naval, Prospero. (2018). Alphabet Sign Language Image Classification Using Deep Learning. 0646-0650. 10.1109/TENCON.2018.8650241.

Adithya V., Rajesh R., A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition, Procedia Computer Science, Vol. 171, 2020, ISSN 1877-0509, https:// doi.org/ 10.1016/ j. procs.2020.04.255.