

Using Deep Convolutional Networks for Sign Language Recognition

May 22, 2019

Nils Thoma, Ramzi Ourari, Achref Doula

1 Proposal

1.1 Abstract

In the context of this project we intend to implement a deep convolutional neural network architecture proposed by[1], propose variances derived from the latter architecture and study the influence of each hyperparameter on the performance of the network. As a bonus we intend to embed our best model into a pc application that translates a sequence of sign language characters into text. This appears to be of a great utility considering that this may present the first step towards a real time translator on smartphones that we intend to implement beyond the scope of this project.

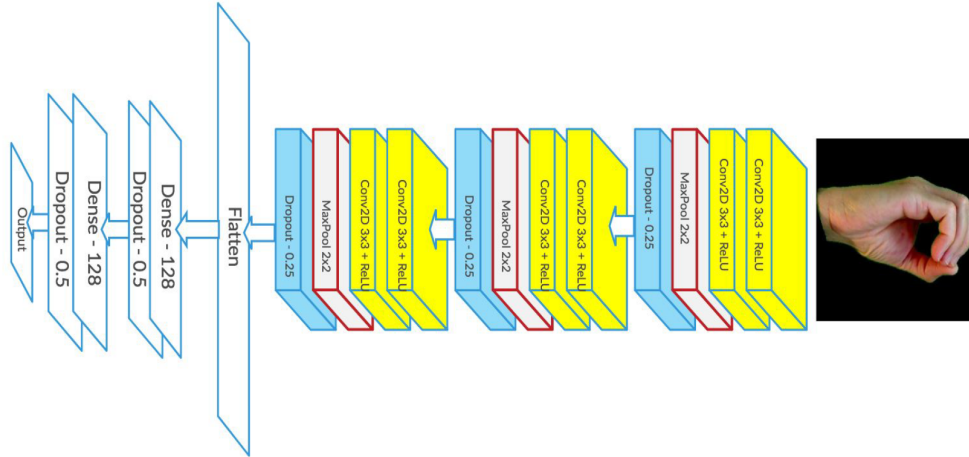
1.2 Motivation

Language translation is one of the core AI tasks. AI systems have reached a competitive performance in the translation between various spoken and written languages. An important form of language, which is well spread but yet not sufficiently treated, is the sign language. More than 460 million people[3] around the world use sign language to communicate with one another or to translate from/to other languages. Automating the task of translating from or to sign language appears then to be a perfect application for Deep Learning solutions. In fact many research institutions and companies have tried to offer deep learning based solutions but as far as we know, no publicly available official paper implementations have been published.

In this project we therefore try then to implement a proposed solution by "Vivek and Dianna" from the "State University of New York at Buffalo" from their paper "Using Deep Convolutional Neural Networks for Gesture Recognition in American Sign Language". Since an official implementation has not been published, we consider implementing the proposed architecture and reproducing the results to familiarize our selves with deep learning frameworks. We then intend to propose other variants of the architecture and study their influence on the network's performance.

This proposal is then description of what we intend to do during this project and contains:

- description of the used Data Set
- details about the baseline architecture
- the set hyperparameters variation that we intend to study



Proposed Network Architecture [1]

1.3 Dataset

The data set used in this project will be the "MUHandImagesASL", made publicly available by the Massey University[2]. The dataset is based on American Sign Language hand gestures. The particularity of this dataset lies on covering large variety of hands, dialects and illumination conditions.

The dataset contains 2425 images of 5 individuals, divided in 5 separated folders, one for each volunteer.

1.4 Reference Paper

1.4.1 Method

For the task of classifying sign Language, Convolutional Neural Networks (CNNs) are used. CNNs have been extremely successful in image recognition and classification problems, and have been successfully implemented for human gesture recognition in recent years [1].

1.4.2 Architecture

The implemented architecture consists of a CNN with multiple convolutional and dense layers followed by a max-pool layer and a dropout layer and two groups of fully connected layer, followed by a drop-out layer and one final output layer

1.4.3 Proposed variations

We will make use of random-search hyper-parameter optimization to determine the best parameters for this architecture and benchmark the results with the published accuracy values. The set of parameters we will cover in this search are:

- Number of layers
- Density of layers
- The drop-out rate

- Learning rate of the optimizer
- The regularization term

Besides, we will compare different optimizers and different loss-functions. Finally, we will compare the effect of using batch-normalization in comparison to setting a dropout-rate

1.5 References

- [1] :Using Deep Convolutional Networks for Gesture Recognition in American Sign Language, Vivek Bheda and N.Dianna Radpour, State University of New York at Buffalo.
- [2] :A. L. C. Barczak, N. H. Reyes, M. Abastillas, A. Piccio, and T. Susnjak, A new 2d static hand gesture colour image dataset for asl gestures, Research Letters in the Information and Mathematical Sciences 15 (2011), 12–20.
- [3] :<https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>