

CS412 Project presentation

Alberto M. Bellini, Matteo Biasielli, Yatri Modi
Simran Jumani, Davide Santambrogio, Federico Sandrelli

1

Department of Computer Science
University of Illinois at Chicago

1. Dataset

The dataset we plan to use is the American Sign Language Dataset (ASL Dataset). The dataset contains about 87000 samples representing images of alphabets in the ASL. The following images provides a brief example of what images in the dataset are like:



Fig. 1. On the left the 26 classes, on the right letter A

Images in the dataset are divided in 29 classes: letters from A to Z, “space”, “backspace” and “nothing”. The presence of the three extra classes “space”, “backspace” and “nothing” is really useful in real time applications.

This dataset also comprises a very tiny 29-images test set. The publisher of the dataset points out that the test set is purposely this small to encourage the use of other external real data.

The dataset is available on Kaggle at: <https://www.kaggle.com/grassknotted/asl-alphabet/home>

2. Task

The task we want to address is mainly the classification of each image into the corresponding class. On top of that, once a good predictive model is created, with the right pre-processing and post-processing we will try to exploit the following tasks:

- Build a real-time stable classifier that takes inputs from a live stream (say a webcam) and correctly recognizes the sign.
- Translate a video with a moving hand into the correct sentence.

3. Techniques we plan to employ

We are planning to compare the performances of the Machine Learning techniques we were taught during the course with some other techniques such as Convolutional Neural Networks and SVMs. The main goal of this project is, other than building an efficient and useful application, understanding which models work better with certain image classification tasks and give an effective comparison of the performances of different models.

Considering how tiny is the test set that is provided, we are also planning to collect some real test data from other external sources or produce the data ourselves (e.g. we can shoot videos and use the labelled frames as test).

In order to try out all the possibilities, we will also consider techniques that are commonly used to push predictive models towards better generalization, such as:

- Dropout (for neural networks).
- Regularization (Lasso and Ridge).
- Data augmentation (apply coloured filters to images, shoot videos with blue screen to be able to change the background, etc.)
- Early stopping (for neural networks)

If necessary, heuristics and models ensembles will be taken into considerations.

4. References

Figure 1 comes from <https://www.nidcd.nih.gov/sites/default/files/Content>