

# Deliverable 1 - Data Selection Proposal

Sheheryar Parvaz

October 5th 2020

## 1 Dataset

The main dataset will be a kaggle dataset of images of the American Sign Language (ASL) alphabet[2]. This dataset contains 87 000 images, with 26 classes for the letters and three more for space, delete, and nothing.

There is also a secondary dataset of recordings of participants saying words in ASL[3]. The task of interpreting sign language in motion is much more complex, and tackling it is only intended as a stretch goal if I have enough time.

## 2 Methodology

### 2.1 Data Preprocessing

The set of images does not need much preprocessing as the data is well organised and the images are clear enough. However, it might be useful to generate more data by applying transformations to the images, such as rotating slightly or adding noise in order to create more relevant data and thus strengthen.

Additionally, it might be useful to analyse the position of the hand and fingers themselves to build a more robust model, though only further testing will tell.

### 2.2 Model

As the task needs to be done in real time, I first need to do object detection to find the hands which are forming the sign. One such model is YOLOv3[1], which is used for real time object detection and can probably be tuned specifically to find hands. Then, I would pass this information to the classification model.

Considering sign language recognition is an image classification problem, a good model to use would be Convolutional Neural Networks (CNN), created and trained using PyTorch.

### 2.3 Evaluation

For this classification problem, all the classes have approximately the same frequency as they are just letters of the alphabet. Thus, accuracy should be a sufficient as a metric. I'm personally aiming for at least 85% accuracy on over 26 classes.

## References

- [1] Redmon, Joseph and Farhadi, Ali. *YOLOv3: An Incremental Improvement*. arXiv:1804.02767, 2018.  
<https://arxiv.org/pdf/1804.02767.pdf>
- [2] Akash. *ASL Alphabet*. Retrieved 2020-10-05.  
<https://www.kaggle.com/grassknotted/asl-alphabet>
- [3] V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, A. Stefan, Q. Yuan and A. Thangali. *The ASL Lexicon Video Dataset*. CVPR 2008 Workshop on Human Communicative Behaviour Analysis, 2008.  
[http://vlm1.uta.edu/~athitsos/asl\\_lexicon/](http://vlm1.uta.edu/~athitsos/asl_lexicon/)